Molecular Dynamics Simulations of Escherichia coli Ammonia Channel AmtB

Nygaard, Thomas Pedersen

Publication date:
2007

Document Version
Publisher's PDF, also known as Version of record

Molecular Dynamics Simulations of *Escherichia coli* Ammonia Channel AmtB

Thomas Pedersen Nygaard

Ph.D. Thesis

Department of Chemistry
Technical University of Denmark

Kgs. Lyngby 2007

Type	Ph.D. thesis
Text and layout	Thomas P. Nygaard
E-mail	tpnygaard@gmail.com
Thesis title	Molecular dynamics simulations of *Escherichia coli* ammonia channel AmtB
Thesis submitted	June 2007 Department of Chemistry Technical University of Denmark (DTU), Kgs. Lyngby, Denmark
Defense & approval	October 2007
Supervisors (DTU)	Günther H. Peters Flemming Y. Hansen Morten Ø. Jensen
Supervisor at Parc Científic de Barcelona	Dr. Carme Rovira
Examiners	Johan Åqvist Birgit Schiøtt Klaus B. Møller
This version	March 2008 Prepared with LaTeX Printed by Frydenberg A/S, Copenhagen, Denmark
ISBN	87-91233-15-1
EAN	9788791233159

Acknowledgment

I would like to express my gratitude to my supervisors Günther H. Peters, Flemming Y. Hansen and Morten Ø. Jensen, who guided this work and was very helpful in the process of writing manuscripts, and Morten, especially, for directing me toward the subject of this thesis; *E.coli* AmtB. From the Department of Chemistry and from the MEMPHYS center I would like to thank Klaus Møller, Ask Jacobsen and Jacob Sonne for valuable discussions and of course thank the MEMPHYS center director Ole Mouritsen for giving me the opportunity to study the *E.coli* protein AmtB at MEMPHYS - a Center of Excellence.

I am also truly thankful to Carme Rovira for granting me the opportunity to visit her group in Barcelona. My daughter, my wife and I, very much enjoyed the time we spent there, and I, the informal atmosphere provided at Centre de Recerca en Química Teòrica. Thanks to all the people at the centre for being welcoming and helpful.

I acknowledge financial support by the Danish National Research Foundation via a grant to the MEMPHYS - Center for Biomembrane Physics, and computing time provided at the Danish Center for Scientific Computing DCSC at the University of Southern Denmark, Odense. I gratefully acknowledge support by HPC-Europa and computing time provided at the Barcelona Supercomputing Center BSC, Barcelona.

Last and most, I am indebted to my wife Bolette, without whom I would not have been able to complete this thesis, and to my daughter Silje - finally I will have more time to play with you.

This version of the thesis has been updated after the Ph.D. defense. Only linguistic problems and misprints have been corrected.

Preface

Solute transport and other kinds of protein function are dynamical processes, where conformational changes, sidechains movements, and charge transfer processes may play important roles. Real experiments are invaluable to probe protein function, but mechanistic insight at a molecular level is not easy to derive therefrom. Although x-ray structures provide structural information at atomic resolution, they are static structures and do not necessarily reveal important structural differences and conformations in the presence or absence of substrate. In many cases, an extension of the structural information into the dynamic domain is therefore crucial for a proper understanding of protein function at the molecular level. Within the last decade technical advances in scientific computing have made it possible to study the atomic scale behavior of large and complex systems such as biomolecules by molecular dynamics (MD) methods. Biomolecular MD simulation has proved to be a powerful tool in the exploration of protein function, not only by itself but also by complementing real experiments; assisting in the interpretation of experiments and guiding future experiments.

For several decades, the ammonia/ammonium transporter (AMT) family of proteins have been a subject for experimental research, but it was not before 2004 the first structure of an AMT protein; *Escherichia coli* AmtB, was solved. This initiated my research on AmtB by use of MD simulations, which is is summarized in this thesis.

This thesis was submitted to the Department of Chemistry at the Technical University of Denmark on June 27, 2007 in partial fulfillment of the requirements for the Degree of Doctor of Philosophy in Chemistry.

Overview of Thesis

Chapter 1, Introduction. This chapter is intended to introduce the reader to *1*) the characteristics of the ammonia/ammonium transporter (AMT) family of proteins, *2*) ammonia/ammonium in human physiology and human Rhesus proteins; members of the AMT family, and *3*) the *Escherichia coli* protein AmtB, which is used as an AMT model protein in this thesis.

Chapter 2, Molecular Dynamics Simulation. This chapter describes classical and *ab initio* molecular dynamics methods with their approximations and limitations.

Chapter 3, Free Energy Calculations. This chapter provides some theoretical background on the calculation of free energy differences and free energy profiles from equilibrium and non-equilibrium measurements. Implementation of a non-equilibrium method to obtain free energy profiles from MD simulations is addressed. This is used in Chapter 5 to reconstruct free energy profiles for NH_3/NH_4^+ transport along the channel axis of *E.coli* AmtB.

Chapter 4, Classical MD Simulation of AmtB. A huge step forward in the understanding of how AMT proteins function was taken recently when x-ray structures of the *E.coli* AMT protein AmtB were reported. Though static x-ray structures reveal the 3-dimensional structure at atomic resolution they do not reveal much about the dynamical aspects of protein function. Thus, the detailed mechanism of how substrate is recruited and transported across the AmtB is yet to elucidated. In this chapter, the available static structural information is extended into the dynamic domain in order to investigate in molecular detail how the protein structure of AmtB couples to its function. As a first attempt to address this coupling of protein dynamics to substrate transport classical molecular dynamics simulation of AmtB were conducted.

Chapter 5, Steered MD Simulation of AmtB. Based on the equilibrium simulations in Chapter 4 a deprotonation mechanism was suggested. The driving force for such a process is the Gibbs free energy change for the acid-base equilibria of NH_4^+ and NH_3. Free energy profiles for NH_4^+ and NH_3 as function of the translocation pathway, i.e., potentials of mean force (PMFs), can be used to estimate this free energy change. This is the subject of the present chapter. Within the framework of steered molecular dynamics (SMD), external forces are applied to steer NH_4^+ and NH_3 into and out of the pore lumen. The theory presented in Chapter 3 is then applied to the SMD simulation results for each substrate and the corresponding PMFs are reconstructed. Based on these PMFs the region where NH_4^+ most likely deprotonates is localized.

Chapter 6, Hybrid QM/MM MD Simulation of AmtB. Classical molecular dynamics simulations can provide insight into the structure/function relationships of proteins but has some short comings. Two of those are that polarization effects are not explicitly accounted for and that chemical reactions cannot take place. Cation-π interactions have been proposed to be important in binding substrate at the extraluminal periplasmic binding site of AmtB and Amm conduction events are inevitably coupled to proton transfer reactions. Substrate recruitment and recognition, as well as testing of the proton transfer mechanism proposed in Chapter 4 are adressed in this chapter. For that purpose hybrid QM/MM MD simulations of AmtB were conducted.

Chapter 7, Conclusion. A general conclusion on the work presented in this thesis.

Contents

Chapter 1

Introduction

1.1 The AMT Family

Ammonia/ammonium (Amm) is an essential growth factor in many domains of life, i.e., bacteria, yeasts, plants and fungi [1], but not in mammals (see 1.2). Since Amm in the uncharged form (NH_3) can passively diffuse across lipid bilayers, actual Amm transport systems were not thought to exist until the mid 1980s where the existence of NH_4^+ transporters/carriers in bacteria was proposed [2, 3]. Subsequently members of this ammonium transporter (AMT) family have been identified in all six kingdoms of life: Plants, animals, protists, fungi, archaebacteria and eubacteria [4–7], suggesting that the process of Amm transport has been conserved throughout evolution. The AMT family encompasses three members: *1*) the Amm transporter proteins (Amt[1]) mainly found in bacteria and plants [4, 6], *2*) the (methyl)-ammonia/ammonium permeases (Mep) mainly found in fungi [5], and *3*) the Rhesus blood group proteins and their associated glycoproteins (Rh) found in animals [8–10]. Some organisms contain multiple AMT proteins from different family members and with both low and high affinities for Amm in order to conduct Amm when present in high and low concentration, respectively. The AMT family proteins have been extensively reviewed in the literature, see, e.g., Refs. [11–15].

The Amt and Mep members of the family are characterized by having 11 trans-membrane helices, while 12 helices are characteristic of Rh proteins. Although the evolutionary path still is unclear [16], it has been hypothesized that the additional helix indicates that the Rh proteins have evolved from the Amt proteins in gram-negative bacteria [11]. These have a characteristic signal peptide that is cleaved off before membrane insertion, and the additional Rh helix could be a result of a simple loss of the cleavage site followed by incorporation of the signal peptide into the membrane.

[1] In this thesis capitalized AMT refers to the entire family, while Amt refers to a member of this family.

The trans-membrane protein AmtB located in the inner-membrane of the gram-negative eubacteria *Escherichia coli* (*E.coli*) was the first x-ray structure of an AMT family protein that was solved [17, 18]. Subsequently the structures of *Archaeoglubus fulgidus* Amt1 and of AmtB-GlnK complexes (see 1.3) have been reported [20–22]. This gained insight into the 3-dimensional structure of AMT proteins, whereof some conserved structural features are shown in Fig. 1.1 and described below using AmtB as an example (for sequence and structural alignments of different AMT family proteins see [7, 12, 17, 18, 20, 23, 24]). The figure also provides a definition of the channel regions described below. Each AMT protein comprises one pore. Water filled depressions in both the periplasmic and the cytoplasmic surfaces (vestibules) lead into the pore. Several backbone carbonyl oxygens and polar residues line the vestibules and attract cations. The bottom of the periplasmic vestibule forms a cationic binding site, which is separated from the pore lumen by two Phenylalanines partly forming a hydrophobic

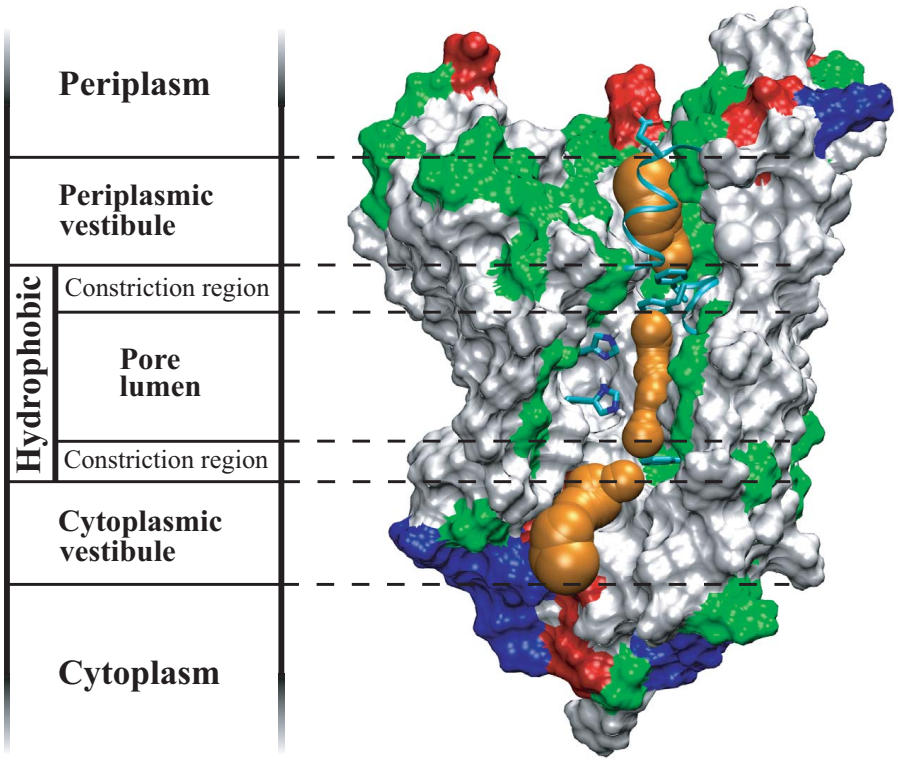

Figure 1.1: The 3-dimensional structure of the AMT family of proteins. Side view of the *E.coli* Amt protein AmtB and definition of channel regions [17, 18]. The protein is shown in surface representation and colored according to residue types; red is acidic, blue is basic, green is polar, and white is nonpolar. Orange spheres represent the pore [19].

constriction region. The pore is an approximately 20 Å long and narrow region, which is formed by conserved, mainly hydrophobic, residues. It is lined by two Histidines in a co-planar arrangement, which share a H-atom between their imidazole rings and which are essential for substrate conductance [25]. Downward, the pore lumen is separated from the cytoplasmic vestibule by yet another hydrophobic constriction region partly formed by one Phenylalanine.

The actual Amm species being transported by AMT proteins and the mechanism of transport have been subjects of debate, e.g., active NH_4^+ transport [4], membrane-potential driven NH_4^+ uniport [6, 26, 27], electroneutral NH_4^+/H^+ antiport [28], facilitated uni- and bidirectional diffusion of NH_3 [29–34], etc. (for a more thorough listing see [35]). Though, the possibility remains that some Amt proteins in plants may facilitate net NH_4^+ transport [13, 35–38], the discussion has otherwise settled on uni- and bidirectional gradient-driven (diffusive) NH_3 conduction [14], not least because of the highly conserved hydrophobic pore lumen revealed by the x-ray structures [17, 18, 20]. The AMT family of proteins should be designated as "diffusive ammonia channels" rather than "ammonium transporters" [14, 17]. Certain AMT proteins are in addition capable of conducting methylammonia (MeAmm), which is a secondary nitrogen source [3, 6, 29, 39, 40], while some Rh proteins have been suggested to conduct CO_2 [30, 41, 42].

Also the Amm species recognized by the AMT proteins was subject of debate. Though NH_3 recognition has been proposed [29, 30] this discussion seems to have settled on NH_4^+ recognition [4, 6, 17, 18, 20, 26–28, 34].

1.2 Ammonia/ammonium in the Human Body

In the human body, Amm is a metabolic waste product mainly produced during amino acid metabolism and primarily in the intestines and in the kidneys [1, 43–45]. Additionally, it is an important participant in systemic acid-base homeostasis [44, 46]. Dysfunctions in the Amm metabolism, e.g., acute liver failure or inherited urea cycle enzyme deficiencies, lead to elevated levels of Amm (hyperammonemia). Hyperammonemia is related to the development of several kinds of brain dysfunctions (encephalopathy), e.g., severe central nervous system dysfunctions, brain edema, herniation, Alzheimer Type II astrocytosis[2] [47–49]. Hence, Amm is highly toxic for human (and mammals in general) and must be kept at low concentration [43]. There are three main routes to "detoxify" Amm: *1*) conversion into Glutamine via Glutamine synthetase mainly in skeletal muscles and in the

[2]Astrocytes are a star-shaped sub-type of the glial cells in the brain. Spanning around the neurons, they form the blood-brain barrier and are thereby important providers of nutrients to the nervous tissue. Astrocytosis is an abnormal increase in the number of astrocytes due to the destruction of nearby neurons.

liver, but also in astrocytes, *2*) direct urinary excretion by the kidneys or *3*) conversion to urea via the ornithine/urea cycle in the liver, [45, 46, 50]. Some of these processes rely on Amm being transported across intracellular compartments and cell membranes. Recently, it was found that the transmembrane Rhesus (Rh) blood group proteins are distantly related to the Amt and Mep proteins and constitute a new member of the AMT family [7, 51]. Before this discovery, the only known function of Rh proteins was that they are involved in cell-surface expression of Rh antigens, and thereby engaged in the well-known immune reactions of red blood cells (erythrocytes). The new discovery suggests that Rh proteins also are clinical important for Amm regulation in human physiology [40, 47]. According to their different localization in cells/tissues, the Rh proteins can be grouped into the erythroid RhAG, RhCE, and RhD proteins and the non-erythroid RhBG and RhCG proteins localized in the liver, kidney, intestinal tract, skin, testis and brain [51]. According to size (weight), the Rh proteins may also be grouped into the nonglycosylated Rh30 proteins (RhCE and RhD; ∼30 kDa) and the glycosylated Rh50 proteins (RhAG, RhBG and RhCG; ∼50 kDa), which also reflects their mutual level of homology. Rh30 proteins in contrast to Rh50 proteins might not be able to promote Amm transport, since they lack the otherwise highly conserved two intraluminal Histidines [25, 52]. This is supported by experiments, showing that Rh50 proteins conduct Amm [33, 36, 40, 53], while no such conduction by Rh30 proteins is observed [25, 33]. So far the exact function of Rh30 proteins as transporters remain unknown. For some Rh50 proteins other experiments have indicated that they might also have a physiological function as CO_2 transporters [41]. For reviews on Amm and Rhesus proteins in relation to the human body see, e.g., Refs. [44, 45, 49, 54].

1.3 *E.coli* AmtB - A Model Protein for Assessing AMT Function

AmtB is the single AMT protein found in *E.coli* [29]. In this organism, Amm is essential for the biosynthesis of Glutamine and Glutamate, which are important starting structures for amino acid synthesis [55]. At low extracellular Amm concentrations *E.coli* growth is AmtB dependent [29], and such nitrogen starvation conditions up regulate the expression of the conserved glnKamtB operon [56–58]. Hence, AmtB is co-transcribed with signal transduction protein GlnK, which is homologous to the P-II protein family of nitrogen regulatory proteins [57]. Intracellular Glutamine pool elevations initiates GlnK deuridylylation as an indirect response on micromolar increase in extracellular ammonia concentration, and deuridylylated GlnK following inactivates AmtB via reversible complex formation. On that background, it has been suggested, that *E.coli* AmtB in coupling

to (de)uridylylation of GlnK constitutes an Amm sensoring system, which plays a role in the intracellular nitrogen regulation mechanism [3, 59–61]. Subsequently, x-ray structures of AmtB-GlnK complexes have been reported [21, 22]. Apparently, the regulation mechanism is conserved in prokaryotes (archaebacteria+eubacteria) [3, 62–66], suggesting that it constitutes an ancestral nitrogen-responsive system [67]. Except for plants, P-II proteins such as GlnK are not found in other eukaryotes (fungi+animals+protists), but the above concept of Amt regulation by cytoplasmic factors makes it conceivable that analogous AMT regulating systems exist in higher organisms as well [11]. For example, the switch of fungi from yeast form to pathogenic filamentous forms occur in response to environmental changes, whereof nitrogen starvation is the switch for some plants and animal (human) pathogens [68, 69].

E.coli is a well-characterized organism with respect to genetics and metabolic pathways. It has therefore been one organism of choice for studying structure/function relationships in AMT proteins in the laboratory [11, 25, 29, 34, 59, 67, 70, 71]. The close relationship between the Amt and Mep members, which also share some of the same structural features important for Amm conduction in the clinical important human Rh50 proteins, suggest that structure/function relationships found in *E.coli* AmtB may also shed light on structure/function relationships in other AMT proteins [12, 17, 23, 33, 52]. Hence, a detailed understanding of how *E.coli* AmtB function has pharmaceutical, agricultural, and biotechnological relevance [15, 68, 72]. Some of the questions which need to be answered are: *1*) How does AmtB recruit substrate? *2*) How does AmtB recognize substrate? And *3*) where and how is NH_4^+ deprotonated?

In this thesis, *E.coli* AmtB is used as a model protein to assess AMT function by molecular dynamics simulation. This structure was an obvious choice being the only available structure at the time this Ph.D. project commenced. The results of my work are presented in Chapter 4, which is adapted from an already published manuscript, Chapter 5, which is based on preliminary results, and Chapter 6, which is adapted from a manuscript in preparation. Since I started working on AmtB, the computational research on AMT proteins, that is *E.coli* AmtB, has accelerated from zero published studies to around eight today [73–80]. To reflect what was known at the time of writing, I found it most appropriate to refer to these studies in the above mentioned chapters rather than in this introduction.

Chapter 2

Molecular Dynamics Simulation

To model molecular systems one may apply the Molecular Dynamics (MD) simulation technique. In an MD simulation the time evolution (dynamics) of the system is modeled by propagating the system according to its equations of motion (e.o.m.), which are solved numerically on a computer. If the system can be properly represented by classical mechanics Newton's formulation of the e.o.m. applies. However, for the purpose of MD simulations of many particle systems it is more convenient to use Lagrange's or Hamilton's formulations. If a classical representation does not apply, the system has to be treated quantum mechanically. In that case the time evolution is obtained by solving the time dependent Schrödinger equation. In the following sections a short introduction to classical MD simulations is provided, as well as to *ab initio* MD simulations. The introductions are not exhaustive and the more interested reader is therefore directed to the vast amount of literature that exists in this field (see e.g. [81–85], and references therein).

2.1 Classical Molecular Dynamics

2.1.1 Hamilton's Equations of Motion

Consider an isolated classical system with N atoms, volume V, and total energy E (NVE). The system is completely specified through the 3N generalized coordinates $\mathbf{q}$ and the 3N generalized momenta $\mathbf{p}$ conjugate to $\mathbf{q}$. The Hamiltonian for the system

$$H(\Gamma) = T_{kin}(\Gamma) + V_{pot}(\mathbf{q}) = E. \tag{2.1}$$

gives the total energy of the system for any given configuration $\Gamma = (\mathbf{q}, \mathbf{p})$, as the sum of the kinetic energy $T_{kin}(\Gamma)$ and the potential energy $V_{pot}(\mathbf{q})$.

The time evolution of the system is governed by Hamilton's e.o.m

$$\dot{p}_i = -\left(\frac{\partial H}{\partial q_i}\right) \qquad \text{and} \qquad \dot{q}_i = \left(\frac{\partial H}{\partial p_i}\right) \quad , \quad i = 1, 2, \ldots, 3\text{N}. \tag{2.2}$$

Inherently, the evolution is deterministic and the total energy is conserved, i.e., $\frac{\mathrm{d}H(\Gamma)}{\mathrm{d}t} = 0$. The e.o.m. may be integrated numerically using finite difference methods, which are based on truncated Taylor series expansions of positions, velocities, etc.. They have different (dis)advantages with respect to energy conservation, time step size, memory usage, time reversibility, phase space volume preservation, etc. (see e.g. [82, 83] and references therein). In an outweighing of these advantages and disadvantages the so-called Verlet-like algorithms have become the method of choice in most MD simulation software. The software used for the work presented in this thesis; NAMD and CPMD (see 2.3), use the velocity-Verlet algorithm [86–88], which easily can be combined with multiple time step algorithms in order to reduce computational costs [89]. The $\dot{p}_i$, in Eq. 2.2, and thereby the force components $f_i = -\left(\frac{\partial V_{pot}}{\partial q_i}\right)$ of the force acting on each particle, has to be evaluated at each discrete time step of the integration. This is in general the most time-consuming part of a classical MD simulation. Consequently, much effort have been made to reduce the computational costs for this part of the calculations by developing/implementing "smart" algorithms and methods, e.g., efficient lists for book keeping purposes [90–92] and efficient ways to treat long range interactions (e.g. electrostatics) without cutoffs in order to avoid truncation errors [93–95].

To propagate the system in time according to Eq. 2.2, the functional forms of $T_{kin}(\Gamma)$ and $V_{pot}(\mathbf{q})$ need to be known. In Cartesian coordinates the kinetic energy is given by

$$T_{kin}(\mathbf{p}) = \tfrac{1}{2} \sum_{i=1}^{3N} \frac{{p_i}^2}{m_i}, \tag{2.3}$$

while the potential energy $V_{pot}(\mathbf{q})$ often is represented by semi-empirical atom-atom potentials (see 2.1.3).

2.1.2 Determination of Thermodynamic Properties

According to the fundamental postulates of classical statistical mechanics it is possible to calculate thermodynamic properties either as an ensemble average or as a time average. The latter allows one to calculate thermodynamic properties from MD simulations. One is often interested in calculating properties corresponding to those measured in real experiments, which normally are carried out at constant temperature T and constant pressure P (or constant volume V). The system considered in the previous

section was subjected to NVE conditions, but other conditions can be realized by applying different constraints, i.e., by coupling the system to baths $(H_1(\Theta_1), H_2(\Theta_2), \ldots)$. For example a heat bath to control the temperature or a volume bath to control the pressure. Let

$$\begin{aligned} H_{tot}(\Gamma, \Theta_1, \Theta_2, \ldots) = H(\Gamma) + H_1(\Theta_1) + H_2(\Theta_2) + \\ \ldots + V_{couple}(\Gamma, \Theta_1, \Theta_2, \ldots) \end{aligned} \tag{2.4}$$

be the Hamiltonian governing the system + baths, where V_{couple} denotes the coupling terms between the system variables (Γ) and the bath variables $(\Theta_1, \Theta_2, \ldots)$. It is not feasible computational wise to couple the system to physical baths, since this would demand that the e.o.m. for the baths have to be solved as well. Instead different algorithms, e.g., the Langevin thermostat and the Langevin piston method [96], can be used to implement temperature and pressure control in an artificial way, which preserves the correct equilibrium ensemble but at a reduced computational costs. These methods are used to ensure constant temperature and pressure in the classical MD simulations conducted in this thesis (see Chapter 4 and 5).

2.1.3 Potential Energy Function and Parameters

A basic functional form of the potential energy $V_{pot}(\mathbf{q})$ in Eq. 2.1 often employed in all-atom MD simulations is

$$\begin{aligned} V_{pot}(\mathbf{q}) &= \overbrace{V_{bond} + V_{angle} + V_{dihedral}}^{\text{bonded}} + \overbrace{V_{vdW} + V_{electrostatic}}^{\text{non-bonded}} \\ &= \sum_{bonds} K^b_{ij}(r_{ij} - r_{eq,ij})^2 \\ &\quad + \sum_{angles} K^a_{ijk}(\theta_{ijk} - \theta_{eq,ijk})^2 \\ &\quad + \sum_{dihedrals} K^d(1 + \cos(n\phi - \delta)) \\ &\quad + \sum_{i<j,\, vdW} 4\epsilon_{ij}\left[\left(\frac{\sigma_{ij}}{r_{ij}}\right)^{12} - \left(\frac{\sigma_{ij}}{r_{ij}}\right)^{6}\right] \\ &\quad + \sum_{i<j,\, elect.} \frac{q_i q_j}{4\pi\epsilon_0\epsilon_r r_{ij}}, \end{aligned} \tag{2.5}$$

where the potential energy function consists of terms, which may be grouped into bonded and non-bonded contributions. The bonded terms explicitly represent bonds, bond angles, and dihedral angles, and only apply to atoms connected through covalent bonds. Except for the dihedral angle term, which has a cosine dependence with phase δ and number of minima n, all other bonded terms are harmonic potentials with force constants K and

equilibrium values of the respective internal coordinates denoted by index *eq*. r_{ij} denotes the distance between atoms i and j, while θ and ϕ denote angles defined by a group of atoms. The non-bonded terms represent long range dispersion interactions (vdW) and electrostatic interactions, and normally apply to all pairs of atoms except those closely connected by covalent bonds; these interactions are either scaled or excluded. The vdW term is usually modeled by a 12-6 Lennard-Jones potential with well depth ϵ_{ij} and collision distance σ_{ij} constructed according to the Lorentz-Berthelot combination rules $\sigma_{ij} = \frac{\sigma_{ii}+\sigma jj}{2}$ and $\epsilon_{ij} = \sqrt{\epsilon_{ii}\epsilon_{jj}}$. The electrostatic term is usually modeled by a Coulomb potential, where q_i and q_j are the partial charges on atoms i and j, and, ϵ_0 and ϵ_r are the vacuum permittivity and the relative dielectric constant, respectively (ϵ_r is normally set to 1, since electrostatic screening effects should be directly accounted for in an all-atom MD simulation). The parameters $(K, r_{eq}, \theta_{eq}, \phi, n, \sigma, \epsilon_{ij}, q)$ entering the different terms are derived from experimental data and quantum mechanical calculations.

The potential energy function $V_{pot}(\mathbf{q})$ together with its parameters is referred to as a molecular mechanical force field. With respect to the basic formula in Eq. 2.5, force fields may differ by the inclusion of varies additional terms such as an improper angle term (to govern planarity of conjugated systems), a hydrogen bond term, an Urey-Bradley term (to harmonically couple atoms separated by two covalent bonds), a cross-term (to couple different bonded terms), etc.. Additionally, the force fields may differ in the way the parameters were derived, e.g., the experimental data used and the applied level of the quantum mechanical calculations. Many different force fields exist and have been developed for different purposes. Two examples of force fields that have been specially optimized for studying biological macromolecules, are CHARMM and AMBER (see 2.3), which both have been utilized in the work presented in this thesis.

Some limitations are inherent to the force fields based on Eq. 2.5. Namely, that covalent bonds due to the harmonic potential cannot "break", which precludes the possibility of chemical reactions to take place. Also, since most force fields treat non-bonded interactions through pairwise potentials and assign a fixed partial charge and fixed Lennard-Jones parameters to each atom type, atoms are unaffected by changes in the local environment. This precludes an explicit account of many-body effects, e.g., (electronic) polarization effects. In the next generation of force fields efforts are made to incorporate models of polarizability [97–99]. Some work has also been done toward the development of reactive force fields [100].

2.2 *Ab Initio* Molecular Dynamics and Hybrid QM/MM

The above limitations of molecular mechanical force fields to account for chemical reactions and polarizability can be avoided by using a quantum mechanical approach instead. Several such approaches to incorporate quantum mechanical methods into MD simulations exist, see e.g. Ref. [88]. Relying on different approximations these so-called *ab initio* MD methods are all derived from the time dependent Schrödinger equation; the quantum mechanical equivalent to the classical e.o.m.. Some approaches rely on the approximation that the nuclei can be treated as classical particles. In that case, one could find inspiration in the classical mechanical approach above and calculate the forces on the nuclei by an electronic structure calculation and subsequently propagate the nuclei according to the e.o.m. in Eq. 2.2. Such a two-step procedure is, however, much more time consuming than conducting the electronic structure calculations on-the-fly with the nuclei propagation. A very efficient approach to conduct *ab initio* MD simulations "on-the-fly" is the Car-Parrinello method [101] (reviewed in [88, 102]). This method is based on the fully *classical* Lagrangian

$$L(\mathbf{q}) = T_{kin}^{nucl}(\dot{\mathbf{q}}) + T_{kin}^{elect}(\dot{\Psi}) - V_{KS}(\mathbf{q}, \Psi) + \sum_{i,j} \Lambda_{ij}(\langle \dot{\psi}_i | \dot{\psi}_j \rangle - \delta_{ij}), \quad (2.6)$$

where $T_{kin}^{nucl}(\dot{\mathbf{q}}) = \frac{1}{2}\sum_n^{3N} m_n \dot{q}_n^2$ is the kinetic energy of the nuclei with mass m, $T_{kin}^{elect}(\dot{\Psi}) = \frac{1}{2}\sum_i^K \mu_i \langle \dot{\psi}_i | \dot{\psi}_i \rangle$ is a fictitious classical kinetic energy term associated with the electronic subsystem with "mass" μ (Ψ is the wavefunction composed of K single electron Kohn-Sham orbitals ψ), $V_{KS}(\mathbf{q}, \Psi)$ is the Kohn-Sham energy (the electronic energy obtained from density functional theory), and the last term is an orthonormality constraint on the orbitals. The corresponding e.o.m. read

$$m_n \ddot{q}_n = -\left(\frac{\partial V_{KS}}{\partial q_n}\right) + \sum_j \Lambda_{ij}\psi_i \quad , \quad n = 1, 2, \ldots, 3N \quad (2.7)$$

$$\mu_i \ddot{\psi}_i = -\left(\frac{\partial V_{KS}}{\partial \psi_i^*}\right) \quad , \quad i = 1, 2, \ldots, K. \quad (2.8)$$

These e.o.m. allow the electrons and nuclei to be evolved simultaneously in real time and at finite temperature. The electronic energy needs only be calculated in the beginning of a simulation, whereafter the e.o.m. ensures that the electrons remain close to the instantaneous ground state during the simulation provided the exchange between the electronic subsystem and the nuclei is small. This decoupling can be ensured by a suitable choice of the fictitious electronic mass.

In comparison with classical MD simulations, *ab initio* MD simulations are computational more expensive. With todays computing power, they are therefore not directly applicable for simulating large biomolecules. However, one may combine the quantum mechanical (QM) approach with the molecular mechanical (MM) force fields described above. Such combinations are called hybrid QM/MM methods (for more details see the recent review by [103] and references therein), and they allow a small part of the system, e.g., an active site, to be treated quantum mechanically, while the remaining part is treated classically. Thus a biomolecule may be studied in its full complexity, but with the computational effort concentrated where it is needed.

2.3 MD Software

Conducting an MD simulation of a large system, e.g., of a biomolecule, is computationally demanding and relies on the use of supercomputers. Though, it is instructive to write the code one self, an efficient implementation of algorithms in a parallel code would be a staggering task. Instead one may choose from highly optimized software packages such as GROMOS [104], NAMD [105], AMBER [106], CHARMM [107], and GROMACS [108], for classical MD simulations, and the latter three as well as CPMD [109–111] for *ab initio* MD simulations. Some MD software comes with its own force field (usually having the same name), whereas other software rely on the use of already existing force fields. In this thesis, NAMD in combination with the CHARMM force field has been used for the purely classical MD simulations presented in Chapters 4 and 5 and CPMD in combination with the AMBER force field for the hybrid QM/MM MD simulations presented in Chapter 6.

2.4 Timescales, System Sizes and Applications

The accessible time scale of an MD simulation is determined by several factors such as the integration time step, the CPU-time expended at each time step, the available computational resources, the system size etc.. The time step used in a classical MD simulation is usually $\sim 10^{-15}$ s (1 fs), which is determined by the frequency of the fastest motion in the system, i.e., the O-H bond stretch. Today, using optimized software on supercomputers to simulate large biomolecular systems (∼100.000–1.000.000 atoms) one can run state-of-the-art classical MD simulations of length ∼100 ns, e.g., the largest all-atom simulation performed today is of the ribosome (2.64 million atoms), which was simulated with NAMD running on 1024 processors yielding 1.5 ns/day [112]. The accessible time scales allow one to study many aspects of protein dynamics except slow processes such as large rearrange-

ments (folding/unfolding) in secondary protein structure, which currently is limited to short polypeptides [113, 114]. However, one may speed up different processes artificially by guiding the system in different directions. There exists several methods to do this. One method is Steered Molecular Dynamics (SMD) where an external force is applied to the system. For example, this method has been used to study the unfolding of short polypeptides [115], and ion and water conduction through channels and nano tubes [116–118]. In Chapter 5, the SMD method is applied for studying NH_3 and NH_4^+ translocation across AmtB.

The time step used in *ab initio* (or hybrid QM/MM) MD simulations is determined by the electronic frequency. The frequency depends on how the electronic subsystem is treated. When using the Car-Parrinello method the electronic frequency is in part governed by the fictitious electronic mass μ. By adjusting μ one may tune the frequency to an optimal low value (allowing one to use a large time step), which at the same time is large enough to assure a decoupling from the nuclei motion. Typical values of μ are 200–1200 a.u. allowing a time step of size 4–8 a.u. (0.1–0.2 fs). *Ab initio* (or hybrid QM/MM) MD methods can be used to study the short time dynamics of systems consisting of ∼100–1000 QM atoms at a picosecond (10^{-12}) time scale, i.e., chemical reactions. QM/MM MD simulations have been successfully applied in the study of biomolecular function where changes in electronic structure play a role, e.g., the enhanced discrimination for CO vs. O_2 in myoglobin in comparison to model compounds [119], drug-DNA interactions of anticancer drugs [120], and a wide range of different enzymatic and catalytic reactions ([121–123]). In Chapter 6, we will use hybrid QM/MM MD simulations to study proton transfer and substrate-protein interactions, i.e., cation-π interactions, in AmtB.

Chapter 3

Free Energy Calculations

In this chapter a few methods to obtain free energy differences and free energy profiles along a reaction coordinate are described. In Chapter 5 one of these methods will be used to reconstruct free energy profiles for transporting NH_3 and NH_4^+ across AmtB. The chapter is organized as follows. First, an introduction to statistical mechanics and nomenclature is provided. Next, methods to obtain free energy differences between equilibrium ensembles by the use of either equilibrium or nonequilibrium measurements are described. Following, the *potential of mean force*; the free energy profile along a given reaction coordinate, is introduced. Finally, methods to reconstruct such profiles by use of either equilibrium or nonequilibrium measurements are described.

3.1 Classical Statistical Mechanics

A system can be described by quantities and potentials like volume V, temperature T, pressure P, number of particles N, etc.. In statistical mechanics is considered an ensemble of systems, which all are constrained to fixed values of some of the quantities/potentials. The most commonly considered ensembles are those of fixed NVT and NPT, which are called the canonical and isobaric-isothermal ensembles, respectively. In the following NVT conditions are assumed, but the extension to the NPT ensemble is trivial. Let $\Gamma = (\mathbf{q}, \mathbf{p})$ define the state of the system, where $\mathbf{q}$ and $\mathbf{p}$ are Cartesian coordinate and momentum vectors, respectively. Let the system be governed by the Hamiltonian $H(\Gamma) = T_{kin}(\mathbf{p}) + V_{pot}(\mathbf{q})$, which for a given state of the system yields the total energy as a sum of the kinetic and potential energy. The potential energy term $V_{pot}(\mathbf{q})$ only depends on particle coordinates, i.e. the spatial configuration of the system, and therefore it is often referred to as the configurational energy. For ideal systems $V_{pot}(\mathbf{q}) \equiv 0$. Let n denote the number of possible states of the system and $H(\Gamma_i)$ denote the energy of

the ith state then

$$Q_{\mathrm{NVT}} = \sum_{i=1}^{n} e^{-\beta H(\Gamma_i)} \sim \int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma)} \tag{3.1}$$

defines the partition function for the NVT ensemble. If the energy levels are close, i.e. in the classical limit, the partition function can be written as an integral instead of a sum [124]. For the sake of simplicity the prefactor h^{-3N} has been excluded in front of the integral. The factor ensures that the classical limit yields the correct value of Q [124]. Likewise, have been excluded the combinatorial prefactor. Normally different symbols are used for the partition function corresponding to different ensembles, here will be used $Q_{\mathrm{NVT}} = Q$ and $\Delta_{\mathrm{NPT}} = \Delta$. For the NPT ensemble the Hamiltonian is added a PV-term, i.e. $\Gamma = (\mathbf{q}, \mathbf{p}, \mathrm{V})$, and correspondingly an integration over volume V must be carried out [125]. For simplicity, the canonical (NVT) ensemble will be considered in the following.

In the canonical ensemble the probability $P(\Gamma_i)$ of being in the ith state is given by

$$P(\Gamma_i) = \frac{e^{-\beta H(\Gamma_i)}}{Q}. \tag{3.2}$$

which is a normalized Boltzmann factor. The ensemble average $\langle \cdots \rangle_{eq}$ of a function f is then given by

$$\langle f \rangle_{eq} = \sum_{i=1}^{n} f \cdot P(\Gamma_i) \sim \int \mathrm{d}\Gamma \; f \cdot P(\Gamma). \tag{3.3}$$

where the index eq denotes that the average is with respect to an equilibrium ensemble. This may seem superfluous, but later path-ensemble averages and non-Boltzmann distributions will also be considered, and hence a distinction is important.

All macroscopic thermodynamic properties can be expressed in terms of the partition functions [126]. Some properties are ensemble averages and can be calculated directly from computer simulations, e.g., Monte Carlo (MC) and Molecular Dynamics (MD) simulations, in the latter case as a time average according to the ergodic hypothesis [82]. Other properties, such as free energies and entropy, are functions of the partition function, which cannot be calculated from simulations. Of special interest here is the free energy at NVT and NPT conditions a.k.a. the Helmholtz and Gibbs free energies, respectively. It appears that these two state functions are particularly simple functionals of the partition functions of the corresponding ensembles

$$A = -\beta^{-1} \ln \, Q \quad \text{and} \quad G = -\beta^{-1} \ln \, \Delta, \tag{3.4}$$

and hence are called the characteristic state functions of the NVT and the NPT ensembles, respectively. When using Cartesian coordinates the independent variables $\mathbf{q}$ and $\mathbf{p}$ can be separated in $H(\Gamma)$. In that case the

partition function factorizes into an ideal and a configurational part

$$Q = Q_{ideal} \cdot Q_{conf} \tag{3.5}$$

The factorization of the partition function allows the free energy to be split into a sum composed of an ideal term and a configurational term. The latter is also referred to as the excess free energy [82]. Often only the configurational part is considered in text book equations, however in this context no such distinction is made.

In the following sections we will consider systems subject to an external constraint, the size of which is governed by an external parameter λ. The entity controlled by λ can be chosen with a lot freedom [127]. For example, λ may govern the size of an external magnetic field, the volume, a constraint on an internal coordinate, the identity of a group of atoms, etc. The Hamiltonian for the constraint system will be denoted by $H(\Gamma;\lambda)$, the equilibrium ensemble sampled for a fixed value of λ is denoted by NVTλ with corresponding partition function $Q(\lambda)$ and Helmholtz free energy $A(\lambda) = -\beta^{-1}\ln Q(\lambda)$.

3.1.1 Work

Consider a process in which the external constraint λ is switched from a value λ_0 to a value λ_1 during time $t = 0 \rightarrow \tau$. By only allowing the parameter λ to couple to the physical system and not to any incorporated baths Eq 2.4 becomes

$$H_{tot}(\Gamma, \Theta_1, \Theta_2, \ldots; \lambda) \approx H(\Gamma;\lambda) + H_1(\Theta_1) + H_2(\Theta_2) + \ldots \tag{3.6}$$

where the approximation sign reflects that coupling terms between the system variables and baths variables have been neglected. For some purposes, such an approximation can be permissible whenever the couplings terms are small compared to the other terms [128]. Of course when deducing the equations of motion for the total system the coupling terms need to be included. In this approximative form the time derivative of H_{tot} is given by

$$\begin{aligned}\frac{\mathrm{d}H_{tot}}{\mathrm{d}t} \approx & \frac{\partial H(\Gamma;\lambda)}{\partial\lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} + \sum_i \frac{\partial H(\Gamma;\lambda)}{\partial\Gamma_i}\frac{\mathrm{d}\Gamma_i}{\mathrm{d}t} \\ & + \sum_i \frac{\partial H_1(\Theta_1)}{\partial\Theta_{1i}}\frac{\mathrm{d}\Theta_{1i}}{\mathrm{d}t} + \sum_i \frac{\partial H_2(\Theta_2)}{\partial\Theta_{2i}}\frac{\mathrm{d}\Theta_{2i}}{\mathrm{d}t} + \ldots\end{aligned} \tag{3.7}$$

Here, Γ_i denotes a single coordinate or momentum, and it should not be confused with the nomenclature used later, where Γ_i denotes a single configuration of the system. For Hamiltonian evolution the above equation is identically zero for fixed λ (cf. Section 2.1). In fact, in this approximate form all terms have be zero, which is an artifact from excluding the coupling

terms. In the exact equation, the individual terms are allowed to vary, and only their sum have to be zero. However, assuming that only H has an explicit λ dependence, then for both cases it follows that

$$\frac{\mathrm{d}H_{tot}}{\mathrm{d}t} = \frac{\partial H(\Gamma;\lambda)}{\partial \lambda}\frac{\mathrm{d}\lambda}{\mathrm{d}t} = \frac{\partial H(\Gamma;\lambda)}{\partial \lambda}\dot{\lambda} \tag{3.8}$$

The change $dH_{tot} = d\mathcal{Q} + d\mathcal{W}$ is a sum of the heat and work exchanged with the surroundings during the process. Since, the total system is thermally isolated then $d\mathcal{Q} = 0$ and following

$$\mathcal{W}(\mathbf{\Gamma}) = \int_{t=0}^{\tau} \mathrm{d}t \frac{\partial H(\Gamma_t;\lambda)}{\partial \lambda}\dot{\lambda}. \tag{3.9}$$

$\mathcal{W}(\mathbf{\Gamma})$ is the external work performed on the *microscopic* system along the path $\mathbf{\Gamma} = \{\Gamma_0, \Gamma_{t_1}, \Gamma_{t_2} \ldots, \Gamma_\tau\}$ of phase space points Γ followed during the process. In this context is used the sign convention that positive work is energy entering the system.

Three different kinds of work can be associated with the process; the reversible work W_r, the external work $\mathcal{W}(\mathbf{\Gamma})$, and the dissipative work $\mathcal{W}_d(\mathbf{\Gamma})$. They are related by

$$W_r = \mathcal{W}(\mathbf{\Gamma}) - \mathcal{W}_d(\mathbf{\Gamma}), \tag{3.10}$$

Per definition $W_r = A(\lambda_1) - A(\lambda_0) = \Delta A$ equals the free energy change between the initial and final state, and hence is path independent. Contrary, $\mathcal{W}_d$ and $\mathcal{W}$ are both path dependent, i.e., path functions. However, if the process is carried out infinitely slowly, thereby ensuring that the system is in equilibrium at all intermediate states, then $\mathcal{W}(\mathbf{\Gamma}^{eq}) = W_r$ and $\mathcal{W}_d(\mathbf{\Gamma}^{eq}) = 0$. Otherwise $\mathcal{W}(\mathbf{\Gamma}^{neq}) \neq W_r$ and $\mathcal{W}_d(\mathbf{\Gamma}^{neq}) \neq 0$. If the same process was carried out in a macroscopic system then $\langle \mathcal{W}(\mathbf{\Gamma}^{neq}) \rangle_F > W_r$ and $\langle \mathcal{W}_d(\mathbf{\Gamma}^{neq}) \rangle_F > 0$, where $\langle \cdots \rangle_F$ is a path-ensemble average (see also Section 3.3.1). The process is denoted by F for consistency with a nomenclature introduced later. It is noted that in contrast to a macroscopic system, the dissipative work associated with a process in a microscopic system can be negative. For macroscopic systems $\langle \mathcal{W}(\mathbf{\Gamma}^{eq}) \rangle_F = W_r$ can therefore be interpreted as the minimum amount of work required to carry out the process, and following $\langle \mathcal{W}_d(\mathbf{\Gamma}^{neq}) \rangle_F$ as the excess amount of work due to frictional forces in the system.

3.2 Free Energy Differences From Equilibrium Measurements

As mentioned above it is normally not possible to calculate partition functions and hence the free energy of a system cannot simply be obtained from

Eqs. 3.4. Luckily, ensemble averages can normally be obtained with good accuracy. The Helmholtz free energy expression can be rewritten as

$$\begin{aligned} A &= \beta^{-1}\ln\left[\frac{1}{Q}\right] \\ &= \beta^{-1}\ln\left[\frac{C\int \mathrm{d}\Gamma\ e^{+\beta H(\Gamma)}e^{-\beta H(\Gamma)}}{Q}\right] \\ &= \beta^{-1}\ln\left[\int \mathrm{d}\Gamma\ e^{+\beta H(\Gamma)}P(\Gamma)\right] + \beta^{-1}\ln\ C \\ &= \beta^{-1}\ln\left\langle e^{+\beta H(\Gamma)}\right\rangle_{eq} + \beta^{-1}\ln\ C. \end{aligned} \tag{3.11}$$

In the second line the normalization constant C ensures that the introduced integral is unity, i.e. that $C\int \mathrm{d}\Gamma = 1$. In this way it appears that by calculation of the ensemble average $\left\langle e^{+\beta H(\Gamma)}\right\rangle_{eq}$ in an equilibrium simulation, A is readily obtained to within a constant $\beta^{-1}\ln\ C$ of the dynamics. In practice, this method only applies to some non-diffusive systems such as solids and glasses [125]. For diffusive systems, such as systems containing liquids, the exponential average in Eq. 3.11 is difficult to evaluate within a finite simulation time, since the important contributions are large positive values of H and these are only rarely encountered according to the Boltzmann factor in Eq. 3.2 [125–127]. However, most often one is interested in the relative free energy change

$$\Delta A = A(\lambda_1) - A(\lambda_0) = -\beta^{-1}\ln\ \left[\frac{Q(\lambda_1)}{Q(\lambda_0)}\right] \tag{3.12}$$

associated with the change of the external parameter λ rather than an absolute free energy A. Methods have been developed to obtain such free energy differences from equilibrium measurements (see, e.g., Refs. [82, 127, 129–131]). Most widely used are the *perturbation* method and the *thermodynamic integration* method [126]. A brief introduction to both methods are given below. Finally, a short description of the umbrella sampling method is provided, since this method is widely used for reconstructing potentials of mean force, i.e., when λ governs a constraint on internal coordinates. The definition and calculation of PMFs will be the subject of following sections.

3.2.1 The Perturbation Method

The free energy difference between the two equilibrium ensembles NVTλ_0 and NVTλ_1 in Eq. 3.12 can be recast as follows [126]

$$\begin{aligned}\Delta A &= -\beta^{-1}\ln\left[\frac{\int \mathrm{d}\Gamma\ e^{-\beta H(\Gamma;\lambda_1)}}{\int \mathrm{d}\Gamma\ e^{-\beta H(\Gamma;\lambda_0)}}\right] \\ &= -\beta^{-1}\ln\left[\frac{\int \mathrm{d}\Gamma\ e^{-\beta[H(\Gamma;\lambda_1)-H(\Gamma;\lambda_0)+H(\Gamma;\lambda_0)]}}{\int \mathrm{d}\Gamma\ e^{-\beta H(\Gamma;\lambda_0)}}\right] \\ &= -\beta^{-1}\ln\left\langle e^{-\beta[H(\Gamma;\lambda_1)-H(\Gamma;\lambda_0)]}\right\rangle_{eq,\,\lambda_0}\end{aligned} \tag{3.13}$$

where in the second line $H(\Gamma;\lambda_0)$ has been added and subtracted in the upper exponential function. Thus, ΔA can now be calculated by generating system configurations according to the NVTλ_0 ensemble and for each configuration Γ calculate the energies $H(\Gamma;\lambda_0)$ and $H(\Gamma;\lambda_1)$. The original derivation was due to Zwanzig [132]. If $H(\Gamma;\lambda_1)$ had been added and subtracted from the lower exponential function, then the following average

$$\Delta A = \beta^{-1}\ln\left\langle e^{\beta[H(\Gamma;\lambda_1)-H(\Gamma;\lambda_0)]}\right\rangle_{eq,\,\lambda_1} \tag{3.14}$$

with respect to ensemble NVTλ_1 would have been obtained instead. In either case the method is numerically accurate only when the regions of phase space occupied by both systems overlap. Essentially, the one Hamiltonian have to be a perturbation of the second one. Hence the method is called the *perturbation* method and the two equations are called the forward and reverse perturbation formulas, respectively [125, 126].

If the initial and final ensembles of the system differ considerably, this is circumvented by designating some path between the two ensembles along intermediate values of λ [82, 87, 125, 126]. If dividing the path into $K-2$ intermediate system states (K sampling windows), then the overall free energy difference can be obtained from a series of perturbation calculations as

$$\Delta A = \sum_{k=0}^{K} -\beta^{-1}\ln\left\langle e^{-\beta[H(\Gamma;\lambda_{k+1})-H(\Gamma;\lambda_k)]}\right\rangle_{eq,\,\lambda_k} \tag{3.15}$$

where $\lambda \in [\lambda_0,\ldots,\lambda_k,\ldots,\lambda_K]$. Here, the final state has been denoted by λ_K instead of λ_1.

The accuracy of the calculation can be addressed by calculating the free energy using both the forward and backward formulas, i.e., for the $\lambda_{0\to K}$ and the $\lambda_{K\to 0}$ process, respectively. The hysteresis of the calculation is a measure of the systematic error [126]. Also, to ensure convergence of the calculations the free energy change within each window should be less than $2k_B$T [126, 127].

3.2.2 The Thermodynamic Integration Method

Differentiation of the Helmholtz free energy $A(\lambda)$ with respect to λ yields

$$\begin{aligned}\frac{\partial}{\partial\lambda}A(\lambda) &= \frac{\partial}{\partial\lambda}\left[-\beta^{-1}\ln\, Q(\lambda)\right] \\ &= \frac{\int \mathrm{d}\Gamma\ \frac{\partial H(\Gamma;\lambda)}{\partial\lambda}e^{-\beta H(\Gamma;\lambda)}}{Q(\lambda)} \\ &= \left\langle \frac{\partial H(\Gamma;\lambda)}{\partial\lambda}\right\rangle_{eq,\,\lambda}\end{aligned} \tag{3.16}$$

Hence by evaluating the derivative of the Hamiltonian for fixed λ-values the free energy difference between ensembles NVTλ_0 and NVTλ_1

$$\Delta A = A(\lambda_1) - A(\lambda_0) = \int_{\lambda_0}^{\lambda_1} \mathrm{d}\lambda \left\langle \frac{\partial H(\Gamma;\lambda)}{\partial\lambda}\right\rangle_{eq,\,\lambda} \tag{3.17}$$

is obtained by integration. This equation was originally derived by Kirkwood [133], and it is known as the *thermodynamic integration* (TI) method [82]. If λ represents a real thermodynamic property of the system it resembles the way free energies are obtained in experiments and hence is termed *natural* TI. Otherwise, it is termed *artificial* TI. Several TI methods exists, which essentially only differ in the way the integrand is evaluated (for details confer e.g. [126]). Contrary, to the perturbation method there is no requirement of overlapping ensembles, instead it is required that $H(\Gamma;\lambda)$ is differentiable [82, 125]. Normally, the linear function

$$H(\Gamma;\lambda) = (1-\lambda)H_I(\Gamma) + \lambda H_{II}(\Gamma) \tag{3.18}$$

in λ is used to couple two systems governed by the Hamiltonians $H_I(\Gamma)$ and $H_{II}(\Gamma)$, and $\lambda \in [0;1]$ is then referred to as a coupling parameter [82]. Using this functional form in Eq. 3.17 yields

$$\Delta A = \int_{\lambda_0}^{\lambda_1} \mathrm{d}\lambda \left\langle H_{II}(\Gamma) - H_I(\Gamma)\right\rangle_{eq,\,\lambda} \tag{3.19}$$

For the above choice of $H(\Gamma;\lambda)$ the following inequality can be derived

$$\frac{\partial^2 A}{\partial\lambda^2} = -\beta\left[\left\langle (V_{II,pot} - V_{I,pot})^2\right\rangle_{eq,\,\lambda} - \left\langle V_{II,pot} - V_{I,pot}\right\rangle^2_{eq,\,\lambda}\right] \leq 0 \tag{3.20}$$

which can be used to check the validity of the TI method [82].

The functional dependence in Eq. 3.18 is widely used for the perturbation method as well. Due to the nature of the nonbonded interaction terms entering the Hamiltonian singularities may arise for $\lambda = 0$ and $\lambda = 1$, which unfortunately can lead to convergence problems [125].

3.2.3 Umbrella Sampling

A prerequisite for using the perturbation method to assess free energy difference between two equilibrium ensembles was that the probability distributions are not too dissimilar or otherwise to designate a path between the two ensembles, such that appropriate similarity between intermediate ensembles is achieved. However, another route to circumvent that problem is to introduce a weighting or biasing function $w(\mathbf{q})$ such that the system samples phase space according to a non-Boltzmann distribution

$$P_U(\Gamma) = \frac{w(\mathbf{q})e^{-\beta H(\Gamma)}}{\int \mathrm{d}\Gamma w(\mathbf{q})e^{-\beta H(\Gamma)}} \tag{3.21}$$

which is related to the unbiased distribution as

$$P(\Gamma) = \frac{P_U(\Gamma)/w(\mathbf{q})}{\langle 1/w(\mathbf{q})\rangle_U} \tag{3.22}$$

This is called the *umbrella sampling* method [134]. The expression in Eq. 3.13 can now be recast into the form

$$\Delta A = -\beta^{-1}\ln\left[\frac{\left\langle e^{-\beta\Delta H(\Gamma)}/w(\mathbf{q})\right\rangle_U}{\langle 1/w(\mathbf{q})\rangle_U}\right] \tag{3.23}$$

where of course the challenge then is to choose a suitable weighting function, which ensures that the relevant (overlapping) regions in phase space are adequately sampled in one single simulation. If that is not possible, then in analogy to the previous procedures the weight function can be made a function of a coupling parameter to sample intermediate states [126].

3.3 Free Energy Differences From Nonequilibrium Measurements

In the previous section were described three methods for calculating free energy differences based on equilibrium measurements. Interestingly, it is also possible to calculate such differences from nonequilibrium measurements, which will be described below.

3.3.1 Path-ensemble Averages - Crooks Theorem

Only few relations are valid for systems which have been driven arbitrarily far away from equilibrium. Crooks have shown that for a thermostated system (NVT) several of these relations can be considered as special cases of the single theorem

$$\langle \mathcal{F}(\mathbf{\Gamma})\rangle_F = \left\langle \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}})e^{-\beta\mathcal{W}_d(\hat{\mathbf{\Gamma}})}\right\rangle_R \quad , \quad \mathcal{F}(\mathbf{\Gamma}) \equiv \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) \tag{3.24}$$

which will be referred to as Crooks theorem [135]. $\mathbf{\Gamma} = \{\Gamma_0, \Gamma_{t_1}, \Gamma_{t_2} \ldots, \Gamma_\tau\}$ denotes a path of phase space points Γ followed in a the *forward* process F during time $t = 0 \to \tau$. The exact reverse path of $\mathbf{\Gamma}$ followed in the *reverse* process R during reversed time $\hat{t} = 0 \to \tau$ is denoted by $\hat{\mathbf{\Gamma}} = \{\hat{\Gamma}_0, \hat{\Gamma}_{\hat{t}_1}, \hat{\Gamma}_{\hat{t}_2} \ldots, \hat{\Gamma}_\tau\}$. $\mathcal{F}(\mathbf{\Gamma})$ and $\hat{\mathcal{F}}(\hat{\mathbf{\Gamma}})$ are arbitrary path functions, and as indicated above they are defined to have the same functional value for identical paths. The path function $\mathcal{W}_d(\hat{\mathbf{\Gamma}})$ is the dissipative work associated with the R process. $\langle \cdots \rangle$ denotes a path-ensemble average over all paths connecting all possible initial and final states, with the restriction that the paths evolve from a system initially in equilibrium (explicit equations are given below in Eq. 3.26). Besides this initial condition there are no restrictions on the time evolution and following the paths are in effect *nonequilibrium* paths. Hence, Crooks theorem relates the path-ensemble average of $\mathcal{F}(\mathbf{\Gamma})$ over a set of forward nonequilibrium paths with the path-ensemble average of $\hat{\mathcal{F}}(\hat{\mathbf{\Gamma}})e^{-\beta\mathcal{W}_d(\hat{\mathbf{\Gamma}})}$ over a set of reverse nonequilibrium paths [135]. The theorem can also be written as

$$\left\langle \mathcal{F}(\mathbf{\Gamma})e^{-\beta\mathcal{W}_d(\mathbf{\Gamma})} \right\rangle_F = \left\langle \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) \right\rangle_R \tag{3.25}$$

where $\mathcal{W}_d(\mathbf{\Gamma}) = -\mathcal{W}_d(\hat{\mathbf{\Gamma}})$. This equation is the starting point for the derivation of the special cases considered in a subsequent section.

The path-ensemble averages $\langle \cdots \rangle$ entering Eqs. 3.24 and 3.25 are explicitly given by

$$\begin{aligned} \langle \mathcal{F}(\mathbf{\Gamma}) \rangle_F &= \sum_{\mathbf{\Gamma}} \rho[\Gamma_0|\beta, \mathbf{E}_0] \cdot \mathcal{P}[\mathbf{\Gamma}|\Gamma_0, \mathbb{M}] \cdot \mathcal{F}(\mathbf{\Gamma}) \\ \langle \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) \rangle_R &= \sum_{\hat{\mathbf{\Gamma}}} \hat{\rho}[\hat{\Gamma}_0|\beta, \hat{\mathbf{E}}_0] \cdot \hat{\mathcal{P}}[\hat{\mathbf{\Gamma}}|\hat{\Gamma}_0, \hat{\mathbb{M}}] \cdot \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) \end{aligned} \tag{3.26}$$

where ρ and $\hat{\rho}$ are the distribution functions for the equilibrium ensembles of initial states Γ_0 and $\hat{\Gamma}_0$ with corresponding state energies $\mathbf{E}_0$ and $\hat{\mathbf{E}}_0$, respectively. Obviously, the distribution functions are related as $\rho[\Gamma_\tau|\beta, \mathbf{E}_\tau] = \hat{\rho}[\hat{\Gamma}_0|\beta, \hat{\mathbf{E}}_0]$. $\mathcal{P}[\mathbf{\Gamma}|\Gamma_0, \mathbb{M}]$ is the probability of following a specific forward path $\mathbf{\Gamma}$ given the initial system configuration is Γ_0 and the dynamics are governed by the vector of transition matrices $\mathbb{M} = \{\underline{\underline{\mathrm{M}}}_0, \underline{\underline{\mathrm{M}}}_{t_1}, \underline{\underline{\mathrm{M}}}_{t_2}, \ldots, \underline{\underline{\mathrm{M}}}_{\tau-1}\}$, whose elements $\mathrm{M}_t\{\Gamma_{t+1}, \Gamma_t\}$ are single timestep transition probabilities. For the reverse process the corresponding entities are denoted as $\hat{\mathcal{P}}$ and $\hat{\mathbb{M}}$. It is a prerequisite that the transition matrices are balanced, which means that the equilibrium distribution $\rho[\Gamma_t|\beta, \mathbf{E}_t]$ at time t is an invariant of $\underline{\underline{\mathrm{M}}}_t$, i.e.,

$$\rho[\Gamma_t|\beta, \mathbf{E}_t] = \underline{\underline{\mathrm{M}}}_t \rho[\Gamma_t|\beta, \mathbf{E}_t]. \tag{3.27}$$

In terms of single time-step transition probabilities $\mathcal{P}[\mathbf{\Gamma}|\Gamma_0, \mathbb{M}]$ can explicitly be written as

$$\mathcal{P}[\mathbf{\Gamma}|\Gamma_0, \mathbb{M}] = \prod_{t=0}^{\tau-1} \mathrm{M}_t\{\Gamma_t, \Gamma_{t+1}\} \tag{3.28}$$

It can be shown that Crooks theorem follows from the condition of *microscopic reversibility* [135], which for an NVT system can be expressed as

$$\frac{\mathcal{P}\left[\mathbf{\Gamma}|\Gamma_0,\mathbb{M}\right]}{\hat{\mathcal{P}}\left[\hat{\mathbf{\Gamma}}|\hat{\Gamma}_0,\hat{\mathbb{M}}\right]} = e^{-\beta\mathcal{Q}(\mathbf{\Gamma})}, \tag{3.29}$$

The path function $\mathcal{Q}(\mathbf{\Gamma})$ is the heat exchanged with the heat bath when following the forward path $\mathbf{\Gamma}$ as described above. A system is microscopic reversible if *i*) its dynamics are Markovian, *ii*) all its state energies are finite, and *iii*) it preserves the equilibrium ensemble when unperturbed. The above expression can readily be generalized to other ensembles, i.e., to the NPT ensemble by multiplying the right hand side by $\exp[-\beta \mathrm{P}\Delta V(\mathbf{\Gamma})]$ where $\Delta V(\mathbf{\Gamma})$ is the volume change associated with the forward process [136]. If not considering an individual path connecting a specific initial state to a specific final state the principle of microscopic reversibility results in detailed balance [137].

Systems governed by Hamiltonian evolution are deterministic and hence Markovian and the equilibrium ensemble is preserved when unperturbed [138, 139]. Implementation of the Langevin dynamics method or the Langevin piston method to simulate isothermal or isothermal-isobaric systems leads to stochastic evolution, which also is Markovian and preserves the equilibrium ensemble [128, 136, 139]. Thus, the condition of microscopic reversibility is met by the classical molecular dynamics simulations conducted in this work (see section 2.1.2), and Crooks theorem applies.

3.3.2 Special Cases

By inserting different functional forms of $\mathcal{F}$ in the path-ensemble average in Eq. 3.24 or Eq. 3.25 one can derive different nonequilibrium relations. By setting $\mathcal{F}(\mathbf{\Gamma}) = \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) = \delta[\beta\mathcal{W}_d' - \beta\mathcal{W}_d(\mathbf{\Gamma})] = \delta[\beta\mathcal{W}_d' + \beta\mathcal{W}_d(\hat{\mathbf{\Gamma}})]$, then Eq. 3.25 yields that

$$\begin{aligned}
\left\langle \delta[\beta\mathcal{W}_d' - \beta\mathcal{W}_d(\mathbf{\Gamma})] e^{-\beta\mathcal{W}_d(\mathbf{\Gamma})} \right\rangle_F &= \left\langle \delta[\beta\mathcal{W}_d' + \beta\mathcal{W}_d(\hat{\mathbf{\Gamma}})] \right\rangle_R \\
\left\langle \delta[\beta\mathcal{W}_d' - \beta\mathcal{W}_d(\mathbf{\Gamma})] \right\rangle_F e^{-\beta\mathcal{W}_d'} &= \left\langle \delta[\beta\mathcal{W}_d' + \beta\mathcal{W}_d(\hat{\mathbf{\Gamma}})] \right\rangle_R \\
P_F(+\beta\mathcal{W}_d') e^{-\beta\mathcal{W}_d'} &= P_R(-\beta\mathcal{W}_d') \\
\frac{P_F(+\beta\mathcal{W}_d')}{P_R(-\beta\mathcal{W}_d')} &= e^{+\beta\mathcal{W}_d'}.
\end{aligned} \tag{3.30}$$

In the second line, it is utilized that the exponential function effectively is a constant factor because of the δ-function and therefore can be moved outside the path-ensemble average. The resulting path-ensemble averages can be regarded as probabilities (in analogy to Eq. 3.46 derived in a later section), i.e., P_F is the probability of following a path where $\mathcal{W}_d(\mathbf{\Gamma}) = \mathcal{W}_d'$

and P_R is the probability of following a path where $\mathcal{W}_d(\hat{\mathbf{\Gamma}}) = -\mathcal{W}'_d$ [135]. In the last line it is assumed that $P_R(-\beta\mathcal{W}'_d) \neq 0$. The result is known as the *transient fluctuation* theorem [135, 136]. The theorem states that the ratio between the forward and backward probabilities of expending an amount of work $\mathcal{W}'_d$ always is equal to $\exp(\beta\mathcal{W}'_d)$.

One may also choose to set $\mathcal{F}(\mathbf{\Gamma}) = \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) = f(\Gamma_\tau) = \delta[\Gamma' - \Gamma_\tau]$, where $\mathcal{F}$ and $\hat{\mathcal{F}}$ now are functions of the final states Γ_τ of the forward process and not the entire path $\mathbf{\Gamma}$. For that choice Eq. 3.25 yields

$$
\begin{aligned}
&\left\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta\mathcal{W}_d(\mathbf{\Gamma})} \right\rangle_F = \left\langle \delta[\Gamma' - \Gamma_\tau] \right\rangle_R \quad \left(= \langle \delta[\Gamma' - \hat{\Gamma}_0] \rangle_R = \langle \delta[\Gamma' - \hat{\Gamma}] \rangle_{eq,\, \hat{t}=0} \right) \\
&\left\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \right\rangle_F e^{\beta\Delta A} = \left\langle \delta[\Gamma' - \Gamma] \right\rangle_{eq,\, t=\tau} \\
&\left\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \right\rangle_F = \frac{\int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma;\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma;0)}} \cdot \frac{\int \mathrm{d}\Gamma \; \delta[\Gamma' - \Gamma] e^{-\beta H(\Gamma;\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma;\tau)}} \\
&\left\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \right\rangle_F = \frac{\int \mathrm{d}\Gamma \; \delta[\Gamma' - \Gamma] e^{-\beta H(\Gamma;\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma;0)}}
\end{aligned}
\tag{3.31}
$$

In going from the first to the second line on the left hand side, it is utilized that $\mathcal{W}_d(\mathbf{\Gamma}) = \mathcal{W}(\mathbf{\Gamma}) - \Delta A$, where $\Delta A = A(\tau) - A(0)$ is defined with respect to the forward process. Since ΔA is a state function, i.e., path-independent, it can be moved outside the path-ensemble average. On the right hand side, the path-ensemble average can be recognized as an ordinary equilibrium ensemble average at time $t = \tau$ by combining Eqs. 3.26–3.28. The remaining steps are trivial manipulations. Integration of the nominator on the righthand side yields that

$$
\begin{aligned}
&\left\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \right\rangle_F = \frac{e^{-\beta H(\Gamma';\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma;0)}} \\
&\int \mathrm{d}\Gamma' \; \left\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \right\rangle_F = \frac{\int \mathrm{d}\Gamma' \; e^{-\beta H(\Gamma';\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma;0)}} \\
&\left\langle e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \int \mathrm{d}\Gamma' \; \delta[\Gamma' - \Gamma_\tau] \right\rangle_F = e^{-\beta(A(\tau) - A(0)} \\
&\left\langle e^{-\beta\mathcal{W}(\mathbf{\Gamma})} \right\rangle_F = e^{-\beta\Delta A}
\end{aligned}
\tag{3.32}
$$

.

In the second line, both sides are integrated with respect to Γ'. Since Γ' is not related to the path $\mathbf{\Gamma}$, the integration can be moved inside the path ensemble average to obtain the third line, where $e^{-\beta\mathcal{W}(\mathbf{\Gamma})}$ becomes a constant factor with respect to this integration. The result is known as the *Jarzynski equality.*

3.3.3 The Jarzynski Equality and Free Energy Differences

Actually, the Jarzynski equation could more easily have been derived by setting $\mathcal{F} = \hat{\mathcal{F}} = 1$ in Eq. 3.25, but the "preliminary" result in Eq. 3.31 will later prove to be useful. The Jarzynski equality was originally derived for isothermal systems [140], but it can readily be extended to isothermal-isobaric systems, essentially by exchanging the Helmholtz free energy A with the Gibbs free energy G [128]. In the above derivation we have considered a so-called forward process which drives the system from an initial equilibrium state at time $t = 0$ to a final out-of-equilibrium state at time $t = \tau$. However, we may freely choose that the process is brought about by changing an external parameter λ, from an initial value λ_0 at time $t = 0$ to a final value λ_1 at time $t = \tau$. Thus, we may rewrite Eq. 3.32 to

$$\left\langle e^{-\beta \mathcal{W}(\mathbf{\Gamma})} \right\rangle_F = e^{-\beta\left(A(\lambda_1) - A(\lambda_0)\right)}, \tag{3.33}$$

and we see that the Jarzynski equality allow us calculate the exact same relative free energy difference in Eq. 3.12 as attempted by the equilibrium methods described in the previous sections. $\mathcal{W}(\mathbf{\Gamma})$ is the external work performed on the system when bringing it from an initial state characterized by λ_0 to a final state characterized by λ_1 and can be calculated from Eq. 3.9.

Since only the initial state, but none of the intermediate states or the final state have to be in equilibrium, the external work entering the equation is in fact the nonequilibrium work. Thereby, the equation interestingly states that from an exponential average of the nonequilibrium work one can obtain the free energy change ΔA between the equilibrium ensembles of the initial and the final states.

For any practical application of Eq. 3.33 the exponential average of the work pose a problem. One can expect that work values $\mathcal{W} = \mathcal{W}(\mathbf{\Gamma})$ measured in a repeated experiment follows some bell-shaped distribution $P(\mathcal{W})$, e.g., a Gaussian distribution

$$P(\mathcal{W}) = \frac{1}{\sqrt{2\pi\sigma_{\mathcal{W}}^2}} e^{-\frac{(\mathcal{W} - \langle \mathcal{W} \rangle)^2}{2\sigma_{\mathcal{W}}^2}}, \tag{3.34}$$

with mean $\langle \mathcal{W} \rangle$ and width $\sigma_{\mathcal{W}}$ (Fig. 3.1). In that case the exponential average can be written as

$$\left\langle e^{-\beta \mathcal{W}} \right\rangle_F = \int \mathrm{d}\mathcal{W}\; e^{-\beta \mathcal{W}} P(\mathcal{W}), \tag{3.35}$$

where the integrand

$$\begin{aligned} e^{-\beta \mathcal{W}} P(\mathcal{W}) &= \frac{N}{\sqrt{2\pi\sigma_{\mathcal{W}}^2}} \exp\left[-\frac{\left(\mathcal{W} - (\langle \mathcal{W} \rangle - \beta\sigma_{\mathcal{W}}^2)\right)^2}{2\sigma_{\mathcal{W}}^2} \right] \\ &= N P_{shift}(\mathcal{W}) \quad , \quad N = \exp\left[\frac{\sigma_{\mathcal{W}}^2 \beta^2}{2} - \beta \langle \mathcal{W} \rangle \right] \end{aligned} \tag{3.36}$$

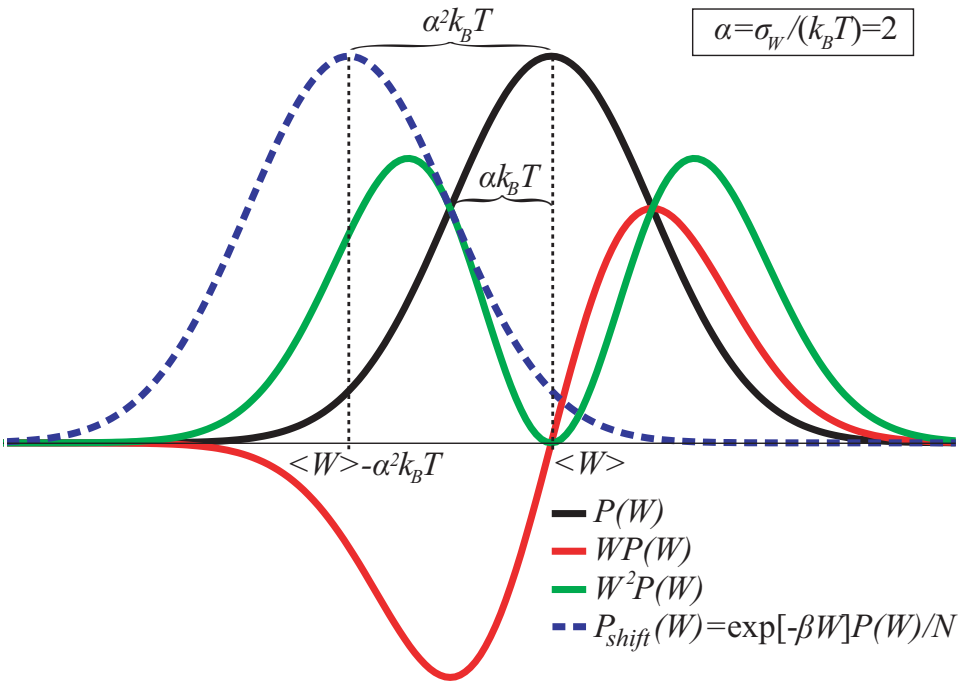

Figure 3.1: Work distributions. Assuming that the work is Gaussian distributed with distribution function $P(\mathcal{W})$, then the integrands entering the different path ensemble averages, which have to be evaluated for a direct application of the Jarzynski equality or when applying a second order cumulant expansion, will also be Gaussian distributed. In contrast to distributions of $\mathcal{W}P(\mathcal{W})$ and $\mathcal{W}^2P(\mathcal{W})$, which are the integrands entering the first and second order moments, the distribution of the integrand entering the exponential average $\exp[-\beta\mathcal{W}]P(\mathcal{W})$ is subject to a $\sigma_{\mathcal{W}}$-dependent shift. For convenience, the shifted distribution $P_{shift}(W) = \exp[-\beta\mathcal{W}]P(\mathcal{W})/N$ is shown. The figure shows the positioning of the respective distribution functions for $\sigma_{\mathcal{W}} = \alpha k_BT$ ($\alpha = 2$). The larger α the more separated is $P_{shift}(W)$ from $P(\mathcal{W})$, and consequently the more difficult it is to evaluate the exponential average of $\mathcal{W}$

also is Gaussian distributed with width $\sigma_{\mathcal{W}}$, but with shifted mean $\langle \mathcal{W} \rangle - \beta\sigma^2_{\mathcal{W}}$. It is this shift which pose a problem, since work values are centered around $\langle \mathcal{W} \rangle$, while an evaluation of the exponential average requires a proper sampling of work values around $\langle \mathcal{W} \rangle - \beta\sigma^2_{\mathcal{W}}$. which are not often encountered [128]. Consequently, many trajectories are needed for a proper evaluation of this slow converging exponential average, which introduces a sampling error on the result. From Eq. 3.36 it is seen that the shift scales as $\sim \sigma^2_{\mathcal{W}}$, while the width only scales as $\sim \sigma_{\mathcal{W}}$. Now, if setting $\sigma_{\mathcal{W}} = \alpha k_B T$ and choosing $\alpha = 2$, then the situation is as schematized in Fig. 3.1. The different scaling behaviors effect that if $\sigma_{\mathcal{W}}$ is larger than a few $k_B T$, as is the case for macroscopic systems, then $P(W)$ will be zero around $\langle \mathcal{W} \rangle - \beta\sigma^2_{\mathcal{W}}$. Thus, in practice a direct application of the Jarzynski equality is limited to microscopic systems, i.e., manipulation of single nanoscale objects [140].

The sampling error can partially be overcome by resorting to the second order cumulant expansion

$$A\big(\lambda_\tau\big) - A(\lambda_0) = \langle \mathcal{W}(\mathbf{\Gamma}) \rangle - \frac{\beta}{2}\left[\big\langle \mathcal{W}(\mathbf{\Gamma})^2 \big\rangle - \langle \mathcal{W}(\mathbf{\Gamma}) \rangle^2\right] \tag{3.37}$$

where $\langle \mathcal{W}(\mathbf{\Gamma}) \rangle$ and $\big\langle \mathcal{W}(\mathbf{\Gamma})^2 \big\rangle$ can be evaluated more accurately, since the distributions of the integrands $\mathcal{W}P(\mathcal{W})$ and $\mathcal{W}^2 P(\mathcal{W})$ both center more closely around $P(\mathcal{W})$ (Fig. 3.1) and are not prone to any $\sigma_{\mathcal{W}}$-dependent shift. Still, $\sigma_{\mathcal{W}}$ should be too large in order for $\langle \mathcal{W}(\mathbf{\Gamma}) \rangle$ and $\big\langle \mathcal{W}(\mathbf{\Gamma})^2 \big\rangle$ to converge fast. The penalty one have to pay by using the faster converging second order cumulant expansion is that the neglection of higher order terms introduces a truncation error to the result [139, 141]. However, when using a stiff spring the work is Gaussian distributed [128], in which case the second order cumulant expansion becomes exact. Comparing the above equation to Eq. 3.10 reveal that for Gaussian distributed work values the following fluctuation-dissipation relation can be established

$$\mathcal{W}_d(\mathbf{\Gamma}) = \frac{\beta}{2}\left[\big\langle \mathcal{W}(\mathbf{\Gamma})^2 \big\rangle - \langle \mathcal{W}(\mathbf{\Gamma}) \rangle^2\right], \tag{3.38}$$

i.e, the dissipative work is related to the width of the work distribution ($\sim$work fluctuation) [140].

3.4 Potential of Mean Force

Again, consider the system sampling its equilibrium ensemble according to the Hamiltonian $H(\Gamma) = H(\mathbf{p}, \mathbf{q})$. The force acting on the ith atom in the system is defined as

$$\mathbf{f}(\mathbf{q}_i) = -\nabla_{\mathbf{q}_i} V_{pot}(\mathbf{q}) = -\nabla_{\mathbf{q}_i} H(\mathbf{p}, \mathbf{q}), \tag{3.39}$$

where $\nabla_{\mathbf{q}_i}$ is the vectorial differiential operator $\left(\frac{\partial}{\partial x_i}, \frac{\partial}{\partial y_i}, \frac{\partial}{\partial z_i}\right)$ with respect to the coordinates $\mathbf{q}_i = (x_i, y_i, z_i)$ of the ith atom. The last equality sign appears since T_{kin} only depends on momenta $\mathbf{p}$ and hence $\nabla_{\mathbf{q}_i} T_{kin}(\mathbf{p}) = 0$.

Now, let $\mathbf{q} = (\mathbf{r}, \mathbf{o})$, where $\mathbf{r}$ are the coordinates of some selected atoms and $\mathbf{o}$ are the coordinates of all other atoms in the system. Assume a structural alteration in the system is well described by the positions of the selected atoms and that they together define a multidimensional reaction coordinate $\mathbf{r}$, i.e., a hyper surface in phase space [125]. Consider the force $\mathbf{f}(\mathbf{r}_i)$ acting on the ith atom comprising the reaction coordinate and the situation where the system is fixed at a given configuration of the reaction coordinate. The average of $\mathbf{f}(\mathbf{r}_i)$ with respect to all possible configurations $\mathbf{o}$ of the other atoms is then given by [124, 126, 142]

$$\begin{aligned}
\langle \mathbf{f}(\mathbf{r}_i) \rangle_{\text{fixed } \mathbf{r}} &= -\langle \nabla_{\mathbf{r}_i} H(\Gamma) \rangle_{\text{fixed } \mathbf{r}} \\
&= \frac{-\int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{o}\ e^{-\beta H(\Gamma)} \nabla_{\mathbf{r}_i} H(\Gamma)}{\int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{o}\ e^{-\beta H(\Gamma)}} \\
&= \frac{\beta^{-1} \nabla_{\mathbf{r}_i} Q(\mathbf{r})}{Q(\mathbf{r})} \\
&= -\nabla_{\mathbf{r}_i} [\underbrace{-\beta^{-1} \ln\, Q(\mathbf{r})}_{=\Upsilon(\mathbf{r})}]
\end{aligned} \tag{3.40}$$

In the second line is seen the effect of fixating $\mathbf{r}$, namely that integration is over $\mathbf{p}$ and $\mathbf{o}$ only. In the third line, has been introduced the partition function

$$Q(\mathbf{r}) = \int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{o}\ e^{-\beta H(\Gamma)} \tag{3.41}$$

for the "fixed $\mathbf{r}$" ensemble and $\nabla_{\mathbf{r}_i}$ is moved outside the integration sign, since the integration is not over $\mathbf{r}$. In the fourth line the characteristic state function

$$\Upsilon(\mathbf{r}) = -\beta^{-1} \ln\, Q(\mathbf{r}) \tag{3.42}$$

for the "fixed $\mathbf{r}$" ensemble is identified, e.g., $\Upsilon(\mathbf{r}) = A(\mathbf{r})$ for the NVT$\mathbf{r}$ ensemble and correspondingly $\Upsilon(\mathbf{r}) = G(\mathbf{r})$ for the NPT$\mathbf{r}$ ensemble. Since minus the gradient of $\Upsilon(\mathbf{r})$ equals the "mean force" acting on any of the reaction coordinate atoms for a given configuration $\mathbf{r}$, it can readily be interpreted as a potential of the "mean force" a.k.a. *potential of mean force* (PMF) [124, 126, 142]. The integral in the partition function $Q(\mathbf{r})$ can be extended to include $\mathbf{r}$ and thereby to encompass all system variables $\Gamma = (\mathbf{p}, \mathbf{o}, \mathbf{r})$ as follows

$$\begin{aligned}
Q(\mathbf{R}) &= \int \mathrm{d}\mathbf{p}\mathrm{d}\mathbf{o}\ e^{-\beta H(\Gamma)} \\
&= \int \mathrm{d}\Gamma\ \left[\prod_{i=1}^{K} \delta(|\mathbf{r}_i - \mathbf{R}_i|)\right] e^{-\beta H(\Gamma)},
\end{aligned} \tag{3.43}$$

In the second line the integrand is multiplied by a product of δ-functions to extract those states where $\mathbf{r} = \mathbf{R}$, thereby ensuring consistency with the first line. The use of δ-functions provides a more simple notation and also allows that the reaction coordinate may be chosen as an arbitrary *function* $\mathbf{r}(\mathbf{q})$ of all system coordinates rather than just representing a *subset* $\mathbf{r}$ of those as above [125]. For the sake of simplicity let us consider situations were the structural alteration of the system can be well characterized by a a one-dimensional reaction coordinate $\mathbf{r}(\mathbf{q}) = r(\mathbf{q})$ [1]. For example it could be the z-component of the center of mass of a substrate when it is being transported through a channel protein with a transport pathway parallel to the z-axis. In that case $r(\mathbf{q})$ will be a mass weighted linear combination of the z-positions of substrate atoms. In the one-dimensional case the expression for the PMF then reads

$$A(R) = -\beta^{-1} \ln Q(R) \tag{3.44}$$

where

$$Q(R) = \int \mathrm{d}\Gamma \; \delta[r(\mathbf{q}) - R] e^{-\beta H(\Gamma)} \tag{3.45}$$

Since $\int \mathrm{d}R \, Q(R) = Q$ the probability distribution function $P(R)$ for finding the system at a given value of the reaction coordinate must be [125]

$$P(R) = \frac{Q(R)}{Q} = \langle \delta[r(\mathbf{q}) - R] \rangle \tag{3.46}$$

Hence the ensemble average of any function $f\big(r(\mathbf{q})\big)$ of the reaction coordinate can be written as

$$\langle f(r(\mathbf{q}) \rangle_{eq} = \int \mathrm{d}R \; P(R) f(R) = Q^{-1} \int \mathrm{d}R \; e^{-\beta A(R)} f(R) \tag{3.47}$$

3.4.1 PMF From Equilibrium Measurements

When combining Eq. 3.44 and Eq. 3.46 the PMF may be written as

$$A(R) = -\beta^{-1} \ln \left[Q \cdot P(R) \right] = -\beta^{-1} \ln P(R) + A \tag{3.48}$$

and thus the PMF $A(R)$ can be determined from the probability distribution of R to within an additive constant, which is the absolute free energy A of the system. If R denotes the distance between two molecules in a homogenous fluid then $P(R)$ is the radial (pair) distribution function, which in some circumstance can be calculated with relatively good precision in a single equilibrium simulation of the system [126]. However, for reaction

[1] In some circumstances r might become a system variable in itself, i.e., by a transformation of system coordinates from a Cartesian to an internal coordinate description [125]. In example, when r represents a distance between two atoms, a torsion angle, or the center of mass of molecule, etc. [127].

coordinates in more complex systems, the statistics obtained from an equilibrium simulation is more likely to be insufficient for reconstructing the PMF [126]. The reason is essentially the same as discussed for Eq. 3.11.

In those cases one may seek to reconstruct the PMF by constraining the system to sample configurations around a given value λ of the reaction coordinate R. In example, equilibrium ensembles for a series of intermediate states along the reaction coordinate can be sampled by using an effective Hamiltonian

$$\mathscr{H}(\Gamma;\lambda) = H(\Gamma) + V\big(r(\mathbf{q});\lambda\big) \tag{3.49}$$

where a biasing potential $V\big(r(\mathbf{q});\lambda\big)$ has been added to enhance sampling of configurations around $r(\mathbf{q}) = \lambda$. Generally, a harmonic potential $V\big(r(\mathbf{q});\lambda\big) = \frac{k}{2}(r(\mathbf{q})-\lambda)^2$ is used. Already at this point one may recognize two problems: *1*) the sought after PMF $A(R)$ is for the unbiased system governed by $H(\Gamma)$ and *not* for the biased system governed by $\mathscr{H}(\Gamma;\lambda)$, and *2*) that the externally controlled parameter is the reference position λ, and not the actual value of the reaction coordinate $r(\mathbf{q})$, which is allowed to fluctuate. Methods to obtain the sought after PMF $A(R)$ of the unbiased system from the equilibrium methods previously described exist [126] (for a comparison of different equilibrium methods see [143]). Widely used is the umbrella sampling method in combination with the weighted histogram analysis method [144]. In that case, the biasing potential applied above corresponds to use a weight function $w(\mathbf{q}) = \exp[-\beta V(r(\mathbf{q});\lambda)]$. Essentially, the PMF can then be obtained from Eq. 3.22 and Eq. 3.48 [126].

3.5 PMF From Nonequilibrium Measurements

In Eq. 3.49 one may also choose to let the reference position λ be time dependent and thereby dynamically guide/steer the system along the reaction coordinate. λ is then called a switching parameter [140]. Normally it is chosen to switch λ linearly in time, i.e., $\lambda(t) = \lambda_0 + vt$, and to apply a harmonic biasing potential

$$V\big(r(\mathbf{q});\lambda(t)\big) = \frac{k}{2}(r(\mathbf{q}) - \lambda(t))^2 = \frac{k}{2}(r(\mathbf{q}) - \lambda_0 - vt)^2. \tag{3.50}$$

This corresponds to the attachment of one end of a Hooke's spring to the molecular assembly defining the reaction coordinate as shown in Fig. 3.2. By moving the other end of the spring along the reaction coordinate an external force is exerted on the molecular assembly, which thereby is forced to move in the same direction. The above process can be realized in atomic force microscope (AFM) or laser tweezers experiments, which allow the manipulation of single molecules, but can also readily be implemented in MD simulations, e.g., steered molecular dynamics (SMD) simulations.

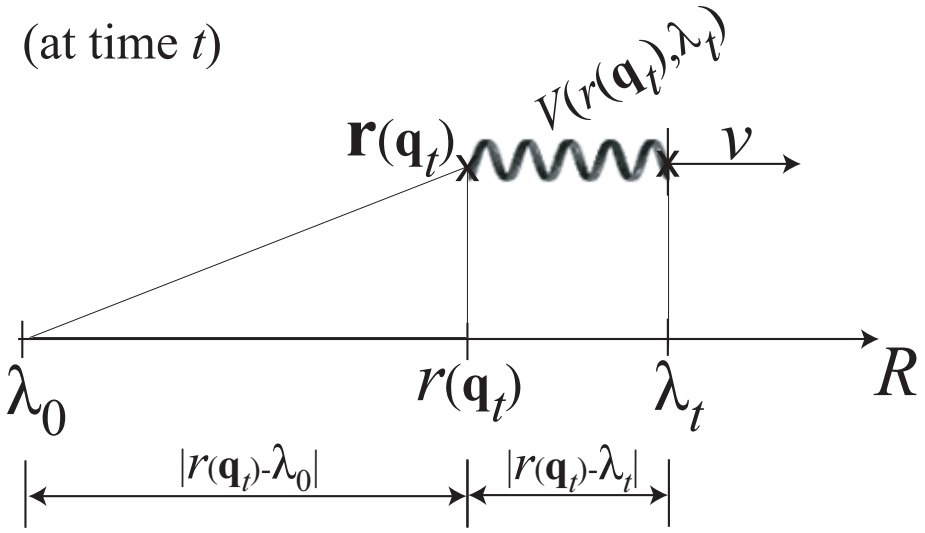

Figure 3.2: Dynamical steering. Schematic drawing of a SMD simulation or an AFM or laser tweezer experiment. λ_0 is the starting value of the reference position $\lambda(t)$, which here changes linearly in time with velocity v, and through an external guiding potential $V(r(\mathbf{q}), \lambda(t)))$ steer a molecular assembly along the reaction coordinate R without applying any constraints perpendicular to the pulling direction. As shown at time t, the actual value of the reaction coordinate $r(\mathbf{q}_t)$ may differ from the reference position λ_t making $V(r(\mathbf{q}_t), \lambda(t))) \neq 0$. For a harmonic guiding potential with force constant k this positional difference affects that an external force $f = -k(r(\mathbf{q}_t) - \lambda_t)$ is exerted on the molecular assembly to guide it towards λ_t.

Obviously, the problems that the sought after PMF $A(R)$ is for the unbiased system and that the reaction coordinate is allowed to fluctuate, also applies here. For reconstructing the sought after PMF $A(R)$, Park et al. and Kosztin et al. have advised two different methods, which both rely on the so-called stiff spring approximation. Within this approximation, Park et al. utilize the Jarzynski equality to obtain the PMF [115], while Kosztin et al. utilize an identity they have derived for forward and reverse pulling experiments; an approach which they all together have named the FR method [145]. The stiff spring approximation is derived from an equation, in which $A(R)$ is expressed *implicitly*. However, it is also possible to reconstruct the PMF based on an equation where $A(R)$ is expressed *explicitly*. Hummer and Szabo have advised such a method, which can be interpreted as a dynamic generalization of umbrella sampling [146]. Here we will refer to this method as the Hummer-Szabo method. Later we will seek to reconstruct the PMF for NH_4^+/NH_3 transport across AmtB using this method, but for completeness all three methods will be described in the following sections.

3.5.1 PMF Reconstruction Using the Stiff Spring Approximation

Consider a time $t = \tau$ corresponding to a value λ_τ of the external control parameter. According to Eq. 3.4 and following paragraphs the equilibrium Helmholtz free energy $\mathscr{A}(\lambda_\tau)$ of the biased system, i.e. the NVTλ_τ ensemble, is

$$\begin{aligned}
\mathscr{A}(\lambda_\tau) &= -\beta^{-1}\ln \int \mathrm{d}\Gamma\ e^{-\beta\mathscr{H}(\Gamma;\lambda_\tau)} \\
&= -\beta^{-1}\ln \left[\frac{Q\int \mathrm{d}\Gamma\ e^{-\beta[H(\Gamma)+V(r(\mathbf{q});\lambda_\tau)]}}{Q}\right] \\
&= -\beta^{-1}\ln \left[Q\left\langle e^{-\beta V(r(\mathbf{q});\lambda_\tau)]}\right\rangle_{eq,\,ub}\right] \\
&= -\beta^{-1}\ln \left[Q\int \mathrm{d}R\ e^{-\beta V(R;\lambda_\tau)}P(R)\right] \\
&= -\beta^{-1}\ln \int \mathrm{d}R\ e^{-\beta V(R;\lambda_\tau)}Q(R) \\
&= -\beta^{-1}\ln \int \mathrm{d}R\ e^{-\beta V(R;\lambda_\tau)}e^{-\beta A(R)}
\end{aligned} \tag{3.51}$$

We note that this equation expresses $A(R)$ *implicitly*. In the second line the integral is multiplied by Q/Q, where $Q = \int \mathrm{d}\Gamma\ e^{-\beta H(\Gamma)}$ is the partition function of the unbiased (ub) ensemble. This allows the introduction of an ensemble average with respect to the unbiased (ub) system in the third line. In the fourth and fifth line the average is written in terms of the reaction

coordinate utilizing the probability distribution function $P(R)$ for the unbiased system introduced in Eqs. 3.46–3.47. In the last line the relation between the partition function and PMF in Eq. 3.44 is used. For a harmonic biasing potential the equilibrium free energy of the biased system at time τ can be recognized as a Gaussian convolution of the PMF $A(R)$ for the unbiased system [115, 128, 145], i.e. the equation can be rewritten as follows

$$\mathscr{A}(\lambda_\tau) = -\beta^{-1}\ln\left[\sqrt{2\pi\sigma^2}\int \mathrm{d}R\; e^{-\beta A(R)}\frac{1}{\sqrt{2\pi\sigma^2}}e^{-\frac{(R-\lambda_\tau)^2}{2\sigma^2}}\right] \tag{3.52}$$

with $\sigma^2 = (\beta k)^{-1}$. It is difficult to obtain $A(R)$ from the above equation, except in the limit of $\sigma^2 \to 0^+$ where the Gaussian distribution function becomes a δ-function, and hence $\mathscr{A}(\lambda_\tau) = A(\lambda_\tau)$ to within a constant. Thus in practice

$$A\left(\lambda_\tau\right) \approx \mathscr{A}(\lambda_\tau) \tag{3.53}$$

if σ^2 is chosen conveniently small, i.e., the force constant k is large ($\sim$ stiff spring).

The problem is now to determine $\mathscr{A}\left(\lambda_\tau\right)$. It is noted that the result is valid for fixed values of λ_τ and hence $\mathscr{A}\left(\lambda_\tau\right)$ can be evaluated by either equilibrium or nonequilibrium methods. Here, we will consider two different approaches to evaluate $\mathscr{A}\left(\lambda_\tau\right)$ from nonequilibrium SMD simulations using either *1*) the Jarzynski equality or *2*) the FR method. Both approaches are introduced in the two following sections.

To check the validity of the approximation $A\left(\lambda_\tau\right) \approx \mathscr{A}(\lambda_\tau)$ higher order correction terms need to be calculated. A Taylor expansion of $\exp[-\beta A(R)]$ in Eq. 3.51–3.52 around λ_τ followed by integration over R leads to the following expression [115]

$$\mathscr{A}(\lambda_\tau) = A\left(\lambda(\tau)\right) - \frac{1}{2k}\left(\frac{\partial A(R)}{\partial R}\right)^2_{R=\lambda_\tau} + \frac{1}{2\beta k}\left(\frac{\partial^2 A(R)}{\partial R^2}\right)_{R=\lambda_\tau} + \mathcal{O}(1/k^2) \tag{3.54}$$

which can be inverted to

$$A\left(\lambda_\tau\right) = \mathscr{A}(\lambda_\tau) + \frac{1}{2k}\left(\frac{\partial \mathscr{A}(\lambda)}{\partial \lambda}\right)^2_{\lambda=\lambda_\tau} - \frac{1}{2\beta k}\left(\frac{\partial^2 \mathscr{A}(\lambda)}{\partial \lambda^2}\right)_{\lambda=\lambda_\tau} + \mathcal{O}(1/k^2) \tag{3.55}$$

and to zeroth order of $1/k$ one may write $A\left(\lambda_\tau\right) = \mathscr{A}(\lambda_\tau)$ as already argued. Hence, the stiff spring approximation has only been a reasonable assumption if the the calculated first order correction term(s) are small compared to the scale of the reconstructed PMF [115]. Otherwise, higher order terms must be included.

Evaluating $\mathscr{A}(\lambda_\tau)$ Using the Jarzynski Equality

$\mathscr{A}(\lambda_\tau)$ can be calculated directly from the Jarzynski equality in Eq. 3.32, which for the biased system reads

$$\left\langle e^{-\beta\mathcal{W}(\mathbf{\Gamma})}\right\rangle_F = e^{-\beta[\mathscr{A}(\lambda_\tau)-\mathscr{A}(\lambda_0)]}. \tag{3.56}$$

The external work $\mathcal{W}(\mathbf{\Gamma})$ entering into this equation, can be calculated from force vs. time curves extracted from, e.g., AFM and laser tweezer experiments. For SMD simulations it can be calculated from Eq. 3.9 as

$$\mathcal{W}(\mathbf{\Gamma}) = \int_{t=0}^{\tau} \mathrm{d}t \frac{\partial \mathscr{H}(\Gamma_t;\lambda)}{\partial\lambda}\dot{\lambda} = \int_{t=0}^{\tau} \mathrm{d}t \frac{\partial V(r(\mathbf{q_t});\lambda)}{\partial\lambda}\dot{\lambda}, \tag{3.57}$$

which for the harmonical biasing potential in Eq. 3.50 becomes

$$\mathcal{W}(\mathbf{\Gamma}) = \int_{t=0}^{\tau} \mathrm{d}t\ vk\big(r(\mathbf{q}_t) - \lambda_0 - vt\big) = \int_{t=0}^{\tau} \mathrm{d}t\ vf(t). \tag{3.58}$$

where $f(t)$ is the force. For practical application of the Jarzynski equality one often resorts to the second order cumulant expansion as previously discussed (see 3.3.3).

Evaluating $\mathscr{A}(\lambda_\tau)$ Using the FR Method

Recently, Kosztin et al. have developed the so-called FR method for calculating the PMF from nonequilibrium measurements [145]. The method is based on a slightly different version of the transient fluctuation theorem in Eq. 3.30

$$\frac{P_F(+\beta\mathcal{W})}{P_R(-\beta\mathcal{W})} = e^{+\beta\mathcal{W}_d}, \tag{3.59}$$

which simply is obtained by setting $\mathcal{F} = \delta(\beta\mathcal{W} - \beta\mathcal{W}[x])$ and $\hat{\mathcal{F}} = \delta(\beta\mathcal{W} + \beta\mathcal{W}[\hat{x}])$, instead. It is assumed that the external work expended in the forward and reverse pulling experiments both are Gaussian distributed, i.e.,

$$P_{F/R}(\beta\mathcal{W}) = (2\pi\sigma^2_{F/R})^{-1/2}\exp\left[-\frac{(\beta\mathcal{W} - \beta\bar{\mathcal{W}}_{F/R})^2}{2\sigma^2_{F/R}}\right] \tag{3.60}$$

where $\bar{\mathcal{W}}_{F/R}$ and $\sigma^2_{F/R} = \sigma^2_{F/R}(\beta\mathcal{W})$ are the mean and variance of the external work in the forward and reverse trajectories, respectively. Plugging the Gaussian distribution functions into Eq. 3.59 and utilizing that $\mathcal{W}_d = \mathcal{W} - \Delta\mathscr{A}$ yields

$$\frac{\sigma_R}{\sigma_F}\cdot\exp\left[-\frac{\beta^2(\mathcal{W}-\bar{\mathcal{W}}_F)^2}{2\sigma_F^2} + \frac{\beta^2(-\mathcal{W}-\bar{\mathcal{W}}_R)^2}{2\sigma_R^2}\right] = \exp\left[\beta(\mathcal{W}-\Delta\mathscr{A})\right]. \tag{3.61}$$

Taking the logarithm and gathering all factors of $\mathcal{W}$ on the left hand side into one term yields

$$\underbrace{\beta\mathcal{W}\left[\underbrace{\frac{\beta\mathcal{W}(\sigma_F^2-\sigma_R^2)}{2\sigma_R^2\sigma_F^2}}_{=0}+\underbrace{\frac{\beta(\sigma_R^2\bar{\mathcal{W}}_F+\sigma_F^2\bar{\mathcal{W}}_R)}{\sigma_R^2\sigma_F^2}}_{=1}\right]}_{=1}-\underbrace{\beta\frac{\sigma_R^2\bar{\mathcal{W}}_F^2-\sigma_F^2\bar{\mathcal{W}}_R^2}{2\sigma_R^2\sigma_F^2}+\ln\frac{\sigma_R}{\sigma_F}}_{\beta\Delta\mathscr{A}}$$
$$=\beta(\mathcal{W}-\Delta\mathscr{A}) \tag{3.62}$$

Comparison of the left hand side to the right hand side reveals that the sum encompassed by the brackets have to be unity. Since the first term contains $\mathcal{W}$ and the second term is constant, the first term must vanish and hence $\sigma_F^2=\sigma_R^2=\sigma^2$. The second term must be unity and hence

$$\beta(\bar{\mathcal{W}}_F+\bar{\mathcal{W}}_R)=\sigma^2 \tag{3.63}$$

Then, it immediately follows that

$$\begin{aligned}\Delta\mathscr{A}&=\frac{\beta(\bar{\mathcal{W}}_F^2-\bar{\mathcal{W}}_R^2)}{2\sigma^2}=\frac{\beta(\bar{\mathcal{W}}_F^2-\bar{\mathcal{W}}_R^2)}{2\beta(\bar{\mathcal{W}}_F+\bar{\mathcal{W}}_R)}=\frac{(\bar{\mathcal{W}}_F-\bar{\mathcal{W}}_R)(\bar{\mathcal{W}}_F+\bar{\mathcal{W}}_R)}{2(\bar{\mathcal{W}}_F+\bar{\mathcal{W}}_R)}\\&=\frac{(\bar{\mathcal{W}}_F-\bar{\mathcal{W}}_R)}{2},\end{aligned} \tag{3.64}$$

i.e, $\Delta\mathscr{A}$ can be evaluated from first order moments of the work and therefore ought to converge faster than the cumulant expansion in Eq. **??**. The second order cumulant expansion additionally depends on the second moment of the work, which is subject to a larger sampling error than the first order moment (Fig. 3.1).

Substituting the above expression for ΔA into the definition of the dissipative work $\bar{\mathcal{W}}_{dF/dR}\equiv\bar{\mathcal{W}}_{F/R}\mp\Delta A$ yield the following relation

$$\bar{\mathcal{W}}_d\equiv\bar{\mathcal{W}}_{dF}=\bar{\mathcal{W}}_{dR}=\frac{(\bar{\mathcal{W}}_F+\bar{\mathcal{W}}_R)}{2} \tag{3.65}$$

and hence the mean dissipative work can be estimated from first order moments alone, as well.

3.5.2 PMF Reconstruction Using the Hummer-Szabo Method

It is important to notice that a direct application of the Jarzynski equality or the FR method to a biased system gives the *exact* free energy change $\mathscr{A}(\lambda_\tau)-\mathscr{A}(\lambda_0)$ for the *biased* system, while only an *approximate* free energy change $A(\lambda_\tau)-A(\lambda_0)$ for the *unbiased* system is obtained. Contrary, the

Hummer-Szabo Method presented below gives the *exact* value for $A(R)$ for the unbiased system.

Hummer and Szabo originally developed the method for reconstructing PMFs from experimental data obtained by micromanipulation of single molecules, e.g., AFM and laser tweezer experiments. The method is based on an equation which can be derived from the first line of Eq. 3.32, which for the biased system reads

$$\Big\langle \delta[\Gamma' - \Gamma_\tau] e^{-\beta \mathcal{W}(\mathbf{\Gamma})} \Big\rangle_F = \frac{e^{-\beta \mathscr{H}(\Gamma';\lambda_\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta \mathscr{H}(\Gamma;\lambda_0)}} = e^{-\beta[\mathscr{H}(\Gamma';\lambda_\tau) - \mathscr{A}(\lambda_0)]}. \quad (3.66)$$

In the original derivation δ-functions was introduced into this equation in a rather arbitrary way [146, 147]. Instead, by going one step back and start from the result obtained in Eq. 3.31 seems more rigorous. The result in Eq. 3.31 was obtained by inserting $\mathcal{F}(\mathbf{\Gamma}) = \hat{\mathcal{F}}(\hat{\mathbf{\Gamma}}) = f(\Gamma_\tau) = \delta[\Gamma' - \Gamma_\tau]$ into Crooks theorem, but we might as well had inserted $\delta[R - r(\mathbf{q}_\tau)]$ instead. With that choice Eq. 3.31 for the biased system reads

$$\begin{aligned}
\Big\langle \delta[R - r(\mathbf{q}_\tau)] e^{-\beta \mathcal{W}(\mathbf{\Gamma})} \Big\rangle_F &= \frac{\int \mathrm{d}\Gamma \; \delta[R - r(\mathbf{q})] e^{-\beta \mathscr{H}(\Gamma;\lambda_\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta \mathscr{H}(\Gamma;\lambda_0)}} \\
&= \frac{Q}{Q} \cdot \frac{\int \mathrm{d}\Gamma \; \delta[R - r(\mathbf{q})] e^{-\beta[H(\Gamma) + V(r(\mathbf{q});\lambda_\tau)]}}{\int \mathrm{d}\Gamma \; e^{-\beta \mathscr{H}(\Gamma;\lambda_0)}} \\
&= \frac{Q \langle \delta[R - r(\mathbf{q})] e^{-\beta V(r(\mathbf{q});\lambda_\tau)} \rangle_{eq,\, ub}}{\int \mathrm{d}\Gamma \; e^{-\beta \mathscr{H}(\Gamma;\lambda_0)}} \\
&= \frac{Q Q^{-1} \int \mathrm{d}R' \; e^{-\beta A(R')} \delta[R - R'] e^{-\beta V(R';\lambda_\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta \mathscr{H}(\Gamma;\lambda_0)}} \\
&= \frac{e^{-\beta A(R)} e^{-\beta V(R;\lambda_\tau)}}{\int \mathrm{d}\Gamma \; e^{-\beta \mathscr{H}(\Gamma;\lambda_0)}}
\end{aligned} \quad (3.67)$$

In the second line the partition function $Q = \int \mathrm{d}\Gamma \; e^{-\beta H(\Gamma)}$ of the unbiased (ub) ensemble is introduced, whereby an unbiased equilibrium ensemble average is obtained in the third line. Utilizing the expression derived in Eq. 3.47 gives the fourth line, which then may be integrated with respect to R' to give the final result. Upon rearrangement it follows that

$$\begin{aligned}
A(R) &= -\beta^{-1} \ln \Big[\Big\langle \delta[R - r(\mathbf{q}_\tau)] e^{-\beta \mathcal{W}(\mathbf{\Gamma})} \Big\rangle_F e^{\beta V(R;\lambda_\tau)} \Big] - \beta^{-1} \ln \int \mathrm{d}\Gamma e^{-\beta \mathscr{H}(\Gamma;\lambda_0)} \\
&= -\beta^{-1} \ln \Big\langle \delta[R - r(\mathbf{q}_\tau)] e^{-\beta \big(\mathcal{W}(\mathbf{\Gamma}) - V(R;\lambda_\tau) \big)} \Big\rangle_F + \mathscr{A}(\lambda_0) \\
&= -\beta^{-1} \ln \Big\langle \delta[R - r(\mathbf{q}_\tau)] e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_\tau)} \Big\rangle_F + \mathscr{A}(\lambda_0),
\end{aligned} \quad (3.68)$$

where

$$\mathscr{W}(\mathbf{\Gamma};\lambda_\tau) = \mathcal{W}(\mathbf{\Gamma}) - V(R;\lambda_\tau).$$

From the above equation the entire PMF $A(R)$ can be obtained *explicitly* to within a constant $\mathscr{A}(\lambda_0)$. The δ-function assures that for each value of R, only the subset of trajectories where $r(\mathbf{q}_\tau) = R$ is used to calculate the average of the "work" $\mathscr{W}(\mathbf{\Gamma};\lambda_\tau)$. $\mathscr{W}(\mathbf{\Gamma};\lambda_\tau)$ is the external work $\mathcal{W}(\mathbf{\Gamma})$ performed up to time τ subtracted the instantaneous biasing potential $V(R;\lambda_\tau)$ at time τ. A direct application of Eq. 3.68 requires infinitely many trajectories, since the reaction coordinate R is a continues variable and additionally has to be sampled several times for each value of R. Therefore, for any practical application one has to coarse grain the reaction coordinate by dividing it into intervals of a finite width Δr. Let R_i be a discrete value denoting the midpoint of the ith interval, and assume that the $r(\mathbf{q}_t)$ values of the subset of trajectories which fall within $R_i \pm \Delta r/2$ are uniformly distributed within this interval, then the δ-function can be approximated by the Boxcar function

$$\begin{aligned}\delta[R - r(\mathbf{q}_t)] &\approx \mathrm{B}_{\Delta r}(R_i - \tfrac{\Delta r}{2}, R_i + \tfrac{\Delta r}{2}) \\ &\approx \Delta r^{-1}[\mathrm{H}(R - R_i + \tfrac{\Delta r}{2}) - \mathrm{H}(R - R_i - \tfrac{\Delta r}{2})].\end{aligned} \tag{3.69}$$

The Boxcar function $\mathrm{B}_{\Delta r}$ is a sum of two Heaviside's stepfunctions H with "steps" located at $R = R_i - \frac{\Delta r}{2}$ and $R = R_i + \frac{\Delta r}{2}$, respectively [148]. Inserting this into Eq. 3.68, and disregarding the additive constant, yields that

$$\begin{aligned}A(R)_{R_i} &\approx -\beta^{-1}\ln\left\langle \mathrm{B}_{\Delta r}(R_i - \tfrac{\Delta r}{2}, R_i + \tfrac{\Delta r}{2}) e^{-\beta\mathscr{W}(\mathbf{\Gamma};\lambda_t)}\right\rangle_F \\ &\approx \beta^{-1}\ln \mathrm{B}_{\Delta r}(R_i - \tfrac{\Delta r}{2}, R_i + \tfrac{\Delta r}{2}) - \beta^{-1}\ln\left\langle e^{-\beta\mathscr{W}(\mathbf{\Gamma};\lambda_t)}\right\rangle_{F,\, R_i \pm \Delta r/2}\end{aligned} \tag{3.70}$$

The two subscripts indicate that the equation only is valid for R values close to R_i and that the average of $e^{-\beta\mathscr{W}(\mathbf{\Gamma};\lambda_t)}$ is over all paths at time τ for which $r(\mathbf{q}_t)$ lies within $R_i \pm \Delta r/2$. The equation is only defined in the interval $R \in]R_i - \frac{\Delta r}{2}, R_i + \frac{\Delta r}{2}[$, where the Boxcar function is unity and hence the first term on the right hand side of Eq. 3.70 vanishes. Outside the interval the Boxcar function is identically zero and the first term leads to infinitely large energies. Obtaining a PMF from the above equation prerequisites that intervals along the entire reaction coordinate R are sampled at time τ, which requires many trajectories. In practice $r(\mathbf{q}_\tau)$ will most likely only sample a part of the reaction coordinate. However, since $\mathcal{W}(\mathbf{\Gamma})$ can be calculated for any intermediate time $t \in [0;\tau]$, it might be possible to sample the entire reaction coordinate by combining the results at different times. In analogy to the umbrella sampling method, a weighted histogram analysis can be used to combine histograms collected at different times, wherefore the method can be regarded as a dynamic generalization of umbrella sampling

[146, 147]. Such an approach still requires many trajectories to obtain a satisfying population of the histograms.

Conveniently, it has been shown that the PMF can be reconstructed from substantially fewer trajectories if a sufficiently large force constant k has been used in the harmonic guiding potential. The important effect of using a relatively stiff spring is that $r(\mathbf{q}_t)$ closely follows $\lambda(t) = \lambda_t$ and hence the fluctuations in $r(\mathbf{q}_t)$ among trajectories at a given time t is small (Fig. 3.3A). This allows one to use all trajectories at time t to estimate the PMF $A(R)$ around λ_t. Additionally, the reconstruction process becomes more simple since the results at different times shall not be combined. Two different approaches to reconstruct the PMF this way have been advised by Hummer and Szabo [147] and Jensen et al. [116]. Differences and similarities between the two approaches will become evident in the following.

For a sufficiently large k one may assume that the positions are Gaussian distributed around a mean position $\langle r_t \rangle$ with variance $\sigma_{r_t}^2 = \langle r_t^2 \rangle - \langle r_t \rangle^2$. Thus, the δ-function should be approximated by a Gaussian distribution function

$$\delta[R - r(\mathbf{q}_t)] \approx (2\pi\sigma_t^2)^{-1/2} e^{-(R-\langle r_t \rangle)^2 (2\sigma_{r_t}^2)^{-1}}, \tag{3.71}$$

rather than a uniform distribution (Boxcar) function as above (Fig. 3.3A). Disregarding the additive constant Eq. 3.68 then becomes

$$\begin{aligned} A(R)_{\lambda_t} &\approx -\beta^{-1} \ln \left\langle (2\pi\sigma_{r_t}^2)^{-1/2} e^{-(R-\langle r_t \rangle)^2 (2\sigma_{r_t}^2)^{-1}} e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_t)} \right\rangle_F \\ &\approx \beta^{-1} \frac{(R - \langle r_t \rangle)^2}{2\sigma_{r_t}^2} + \beta^{-1} \ln (2\pi\sigma_t^2)^{1/2} - \beta^{-1} \ln \left\langle e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_t)} \right\rangle_F \end{aligned} \tag{3.72}$$

where the average of $e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_t)}$ is over all paths at time t. In contrast to the Boxcar approximation, $A(R)$ is defined for all R, though it is only valid for R values close to λ_t, as indicated by the subscript. If we just seek to estimate the value of $A(R)$ for $R = \langle r_t \rangle$ and assume that σ_t is time independent, whereby it just becomes an additive constant to $A(R)$, then Eq. 3.72 reduces to

$$A(\langle r_t \rangle)_{\lambda_t} \approx -\beta^{-1} \ln \left\langle e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_t)} \right\rangle_F \tag{3.73}$$

It is noted that this result resembles that in Eq. 3.70, where a uniform distribution was assumed. In the limit of an infinitely large force constant k then $\langle r_t \rangle \to \lambda_t$ and thereby $V(\langle r_t \rangle, \lambda_t) = 0$. Then the right hand side equals $\mathscr{A}(\lambda_t)$ and hence the stiff spring approximation in Eq. 3.53 is recovered.

The differences between the two different approaches of Hummer and Szabo and Jensen et al. lie in the choice of $\langle r_t \rangle$ and whether Eq. 3.72 or Eq. 3.73 is utilized. Hummer and Szabo utilize Eq. 3.72 and choose

$$\langle r_t \rangle = \langle r(\mathbf{q}_t) \rangle_{eq,\,b} = \frac{\langle r(\mathbf{q}_t) e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_t)} \rangle_F}{\langle e^{-\beta \mathscr{W}(\mathbf{\Gamma};\lambda_t)} \rangle_F} = \frac{\langle r(\mathbf{q}_t) e^{-\beta \mathcal{W}_d(\mathbf{\Gamma})} \rangle_F}{\langle e^{-\beta \mathcal{W}_d(\mathbf{\Gamma})} \rangle_F} \tag{3.74}$$

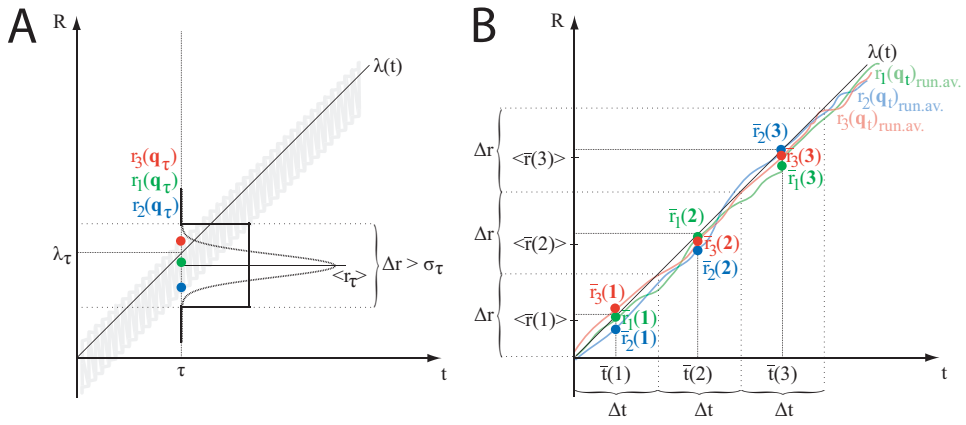

Figure 3.3: δ-function approximations and time averaging. (A) For a relatively large force constant k trajectories follow closely the reference position $\lambda(t)$, but are subject to thermal fluctuations (*light-gray*). At the sharp time τ the values of the reaction coordinate $r_i(\mathbf{q}_\tau)$ encountered in trajectories $i = 1, 2, 3$ (*green, blue, red*) are Gaussian distributed with mean value $\langle r_\tau \rangle$ and variance σ_τ^2. Note, that the Gaussian distribution might be displaced with respect to λ_τ. The reaction coordinate R can be coarse grained in intervals of size Δr chosen larger than the size of the thermal fluctuations. (B) Thermal fluctuations can be filtered out (see 3.5.3), for example by applying a running average to the time series $r_i(\mathbf{q}_t)$ the average $\overline{r}_i(\mathbf{q}_\tau)_{\text{run.av}}$ (*light-colored*) is obtained. At average time $\overline{t}(j)$ in the jth time window, the time averaged reaction coordinate values $\overline{r}_i(j)$ are Gaussian distributed with mean value $\langle \overline{r}(j) \rangle$ and a variance inversely proportional to the width of the running average (see Eq. 3.82), which is smaller than that in A.

which is the equilibrium (eq) average position in the biased[2] (b) system at time t [147]. Jensen et al. utilize Eq. 3.73 and choose

$$\langle r_t \rangle = \langle r(\mathbf{q}_t) \rangle_{neq,\,b} = \langle r(\mathbf{q}_t) \rangle_F \tag{3.75}$$

which is the nonequilibrium (neq) average position in the biased system at time t. This $\langle r_t \rangle$ is easier to calculate, since no "unbiasing" procedure needs to be applied, but instead is obtained as a simple average over the nonequilibrium trajectories. As in Eq. 3.37 the convergence of the exponential averages in Eq. 3.72 and Eq. 3.73 can be improved by using a second order cumulant expansion when only few trajectories are at hand.

3.5.3 Implementation

To reconstruct the PMF associated with translocation of NH_3/NH_4^+ across AmtB, we have chosen to use the Hummer-Szabo method by applying the

[2]Note, the analogy to Eq. 3.22 (umbrella sampling), where the weighting factor $w(\mathbf{q}_t) = e^{+\beta \mathcal{W}_d(\mathbf{\Gamma})}$, in this case is a function of the dissipative work. Here, the "unbiasing" procedure relates *nonequilibrium* to *equilibrium* properties of the *biased system*, which is different from the equilibrium umbrella sampling, where the unbiasing procedure relates *equilibrium* properties of the *biased* system to *equilibrium* properties of the *unbiased* system.

approximation advised by Jensen et al. [116] to the results of SMD simulations. Each MD trajectory provides a practically continuous time series of the reaction coordinate $r(\mathbf{q}_t)$ as visualized in Fig. 3.3A, and of the applied force $f(t)$. From the latter a continues time series of the external work $\mathcal{W}(t)$ is readily obtained from Eq. 3.58. Essentially, the approximation in Eq. 3.73 tells us that $A(R)$ at $R = \langle r_\tau \rangle$ can be obtained by averaging over all paths at time τ. However, it is demanded that $r(\mathbf{q}_t)$ of all trajectories fall within an interval of size Δr and following this interval should be chosen larger than the thermal fluctuations of $r(\mathbf{q}_t)$ (Fig. 3.3A). For a relatively large force constant k the size of the thermal fluctuations is governed by the harmonic biasing potential. Thus, having chosen either k or Δr will restrain the choice of the other parameter. In the following we will consider that Δr has been chosen.

How to Choose the Force Constant k

For a given reference position λ_t the probability of encountering a value R of the reaction coordinate, when using a relatively large force constant, is proportional to the Boltzmann factor

$$P(R - \lambda_t) \propto e^{-\beta \frac{k}{2}(R-\lambda_t)^2} \tag{3.76}$$

which essentially is a Gaussian distribution

$$P(R - \lambda_t) \propto \frac{1}{\sqrt{2\pi\sigma_R^2}} e^{-\frac{(R-\lambda_t)^2}{2\sigma_R^2}} \tag{3.77}$$

where $\sigma_R = (\beta k)^{-1/2}$ is the order of the thermal fluctuations in the reaction coordinate, while the order of the corresponding force fluctuations are $\sigma_f = k\sigma_R = (k/\beta)^{1/2}$ [149]. Since it is demanded that $\Delta r > \sigma = (\beta k)^{-1/2}$, the following requirement to k can be established

$$k > \frac{1}{\beta(\Delta r)^2} \tag{3.78}$$

Another reasonable requirement for k is that it should at least allow the reconstruction of the largest barrier $E_a = E_{max} - E_{min}$ in the PMF whenever E_{min} and E_{max} are separated by a distance larger than the chosen resolution. Demanding that $\frac{1}{2}k(\Delta r)^2 > E_a$ ensures that the global minimum of the biased PMF is always located within $\lambda_t \pm \Delta r$, which leads to the second requirement

$$k > \frac{2E_a}{(\Delta r)^2} \tag{3.79}$$

for the choice of k [145]. E_a might be known in advance, e.g., from experiments.

Time Averaging

There is no point in reconstructing the PMF with a spatial resolution Δr higher than the size of the thermal fluctuations. If the reference position λ_t changes linearly in time with a constant velocity v, then a time window of size $\Delta t = \frac{\Delta r}{v}$ has to pass before λ_t has incremented by Δr (Fig. 3.3A–B). Hence, Δt gives the spacing between successive time points t_j and t_{j+1} where Eq. 3.73 should be applied.

However, a finer spatial resolution $\Delta r' < \Delta r$ can be obtained by filtering out the thermal fluctuations in the time series through some time averaging procedure before averaging over trajectories. Assume that for the ith trajectory $X_{i,t} = r_i(\mathbf{q}_t)$ is a stochastic variable, which at the sharp time t is Gaussian distributed with time dependent mean $\mu_t \approx \lambda_0 + vt$ and time independent variance $\sigma^2 = \sigma_t^2$. We may write $X_{i,t} = r_i(\mathbf{q}_t) - vt + vt = Y_{i,t} + vt$, where the stochastic variable $Y_{i,t} = r_i(\mathbf{q}_t) - vt$ is Gaussian distributed with both time independent mean $\mu \approx \lambda_0$ and variance σ^2. The time averaged value of $X_{i,t}$ over the jth time window of width $\Delta t'$ is given by

$$\begin{aligned} \overline{X}_i(j) &= \frac{1}{\Delta t'} \int_{t_j - \frac{\Delta t'}{2}}^{t_j + \frac{\Delta t'}{2}} \mathrm{d}t \; X_{i,t} \\ &= \overline{Y}_i(j) + v\bar{t}(j) \end{aligned} \tag{3.80}$$

where $\overline{Y}_i(j)$ and $\bar{t}(j)$ are defined the same way as $\overline{X}_i(j)$. Anticipating that $\overline{Y}_i(j)$ and $v\bar{t}(j)$ are stochastic independent the variance of the $\overline{X}_i(j)$ values obtained for different realizations (trajectories) can be split into two terms

$$V(\overline{X}(j)) = V(\overline{Y}(j)) + V(v\bar{t}(j)) \tag{3.81}$$

Since $\bar{t}(j)$ does not vary between trajectories $V(v\bar{t}(j)) = 0$. From statistics it is known that

$$V(\overline{X}(j)) = V(\overline{Y}(j)) = \frac{\sigma^2}{n} \propto \frac{\sigma^2}{\Delta t'} \tag{3.82}$$

where n is the number of time points included in the average [150]. The above n dependence only apply when the $Y_{i,t}$ values at different time points are stochastically independent, which cannot be expected if values for each integration time step in the ith trajectory is used. However, this can be overcome by using time points separated by a sufficient number of integration time steps, and following it results that the variance of $\overline{X}_i(j)$ will be inversely proportional to the width of the time window. Thus, the effect of performing a time average for each trajectory is that the variance of the time averaged values is smaller than the variance of the sharp time values (Fig. 3.3A–B). This allows a spatial resolution $\Delta r'$ which is finer than Δr. Following the approach of Jensen et al. a linear filter is applied to filter out the thermal fluctuations. In principle, a running average of width $\Delta t'$ could

be applied, however, since we only need one value within each subsequent non-overlapping time window, an average over each time window is sufficient. Now, $\Delta t'$ should be chosen such that it encompasses several periods of the fluctuation. The period τ of the harmonic potential is given by

$$\tau = \frac{2\pi}{\sqrt{k/m}} \tag{3.83}$$

where m is the mass attached to the spring. If $\Delta t'$ is chosen too short, i.e., $\Delta t' < \tau$ then the variances on $\overline{X}$ will be of the same size as the thermal fluctuations σ^2.

Time averaged values[3] of $r_i(\mathbf{q}_t)$, $\mathcal{W}_i(t)$, and λ_t, can be calculated from Eq. 3.80. It may be chosen to assign the resulting values $\overline{r}_i(j)$, $\overline{\mathcal{W}}_i(j)$, and $\overline{\lambda}(j) = \lambda_0 + v\bar{t}(j)$ to the time point $\bar{t}(j)$. In terms of time averaged values Eq. 3.68 then reads

$$A(R) = -\beta^{-1}\ln\left\langle \delta[R - \overline{r}_i(j))]e^{-\beta\overline{\mathscr{W}}_i(j)}\right\rangle_F + \mathscr{A}(\lambda_0) \tag{3.84}$$

where

$$\overline{\mathscr{W}}_i(j) \equiv \overline{\mathcal{W}}_i(j) - \overline{V}(j) \quad , \quad \overline{V}(j) = V(R;\overline{\lambda}(j)).$$

and the time averaged version of Eq. 3.73 then becomes

$$A(\langle\overline{r}(j)\rangle)_{\overline{\lambda}(j)} \approx -\beta^{-1}\ln\left\langle e^{-\beta\overline{\mathscr{W}}_i(j)}\right\rangle_F, \tag{3.85}$$

It should be noted that there is a choice in how to introduce the time averages. For example, one might as well have set $\overline{V}(j) = \overline{V}(R;\lambda_t)$. The two choices are related as $\overline{V}(R;\lambda_t) = V(R;\overline{\lambda}(j)) - \sigma^2_{\lambda_t}$. When only considering that λ_t changes linearly with time, then the variance of λ_t will have the same value for each time window. In that case, it effectively just becomes an additive constant to $A(R)$, and the two choices are therefore equivalent.

[3]Essentially, the time averaged values correspond to data points that would be measured in an experiment, where the measuring apparatus probe the system in time intervals of size $\Delta t'$.

Chapter 4

Classical MD Simulation of AmtB

Adapted from [75]: Ammonium Recruitment and Ammonia Transport by *E.coli* Ammonia Channel AmtB. Nygaard, T. P., C. Rovira, G. H. Peters, and M. Ø. Jensen. *Biophysical Journal* (2006) **91**:4401–4412.

4.1 Introduction

Ammonia/ammonium (Amm) is an essential growth factor in bacteria, yeasts, plants, and fungi, while it is toxic to mammalian cells. Proteins belonging to the Amm transporter family have been identified in a wide range of organisms; Amm transporter (Amt) proteins in bacteria and plants, (methyl-)-ammonia/ammonium permeases in yeasts, and Rhesus blood group proteins in animals [8, 9]. Certain Amt proteins, are in addition, capable of transporting methylammonia/-ammonium (MeAmm) [3, 6, 29, 39, 40], while some Rhesus proteins also can transport CO_2 [41].

The first Amt protein for which high resolution crystallographic structures have been solved is *Escherichia coli* Amm transporter B (AmtB) [17, 18]. In one study, three sequentially identical mutated structures were reported [17]. We refer to these as the Khademi structures. A second study reported two wild type (wt) structures [18]. We refer to these as the Zheng structures.

As shown in Fig. 4.1A three identical AmtB monomers form a trimer [17, 18, 70, 71]. The AmtB monomer has 11 transmembrane-spanning α-helices (TM1–TM11) arranged in a right-handed bundle (Fig. 4.1A). Depressions in both the periplasmic and the cytoplasmic surfaces (vestibules) lead into a hydrophobic pore lined by Histidines H168 and H318 (Figs. 4.1B and 4.4 *A*), which, by sharing a H-atom between their imidazole rings, are mutually fixed [17, 18]. Aromatic residues F103, F107, and W148, together with the hydrogen acceptor S219:O_γ, have been suggested to define a substrate

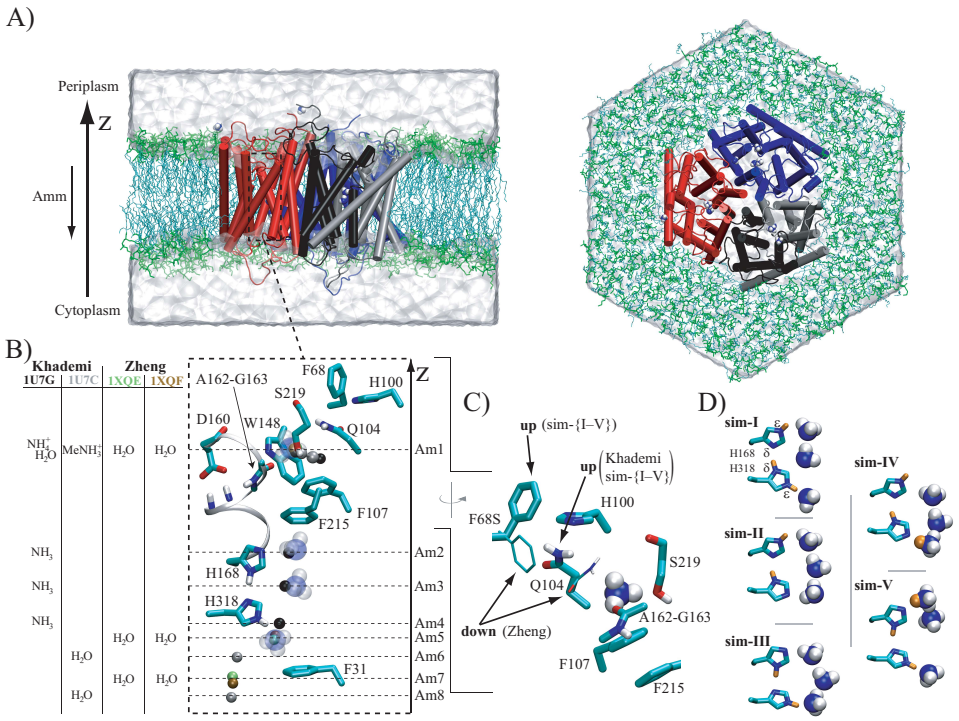

Figure 4.1: The simulated system. (*A*) Snapshot of sim-III (see Table 4.1) showing side and top views of the AmtB trimer with monomers colored red, blue, and gray/black, respectively. Lipid headgroups and hydrophobic tails are colored green and cyan, respectively. Water is depicted in transparent surface representation, and Amm molecules appear in blue and white vdW representation. The quasi twofold symmetry-related helices of one of the monomers are colored gray and black. The physiological direction of Amm transport in *E.coli* is indicated. (*B*) Enlargement of key residues of the substrate binding site; F68, H100, Q104, F107, W148, D160, A162-G163, and S219, and of the pore lumen; F31, H168, F215, and H318 (note the helix-stabilizing function of D160). The corresponding simulation snapshot of NH_3 and NH_4^+ is depicted as transparent vdW spheres. Solid spheres represent selected water molecules, Amm, or MeAmm molecules in the crystal structures of AmtB as tabulated, and their positions are numbered Am1–Am8 (an extension of the numbering scheme used by Khademi et al. [17]). Substrates of the Khademi structures are colored black (1U7G) and gray (1U7C) [17], while those of the Zheng structures are colored green (1XQE) and brown (1XQF) [18]. The positions of water in the third reported Khademi structure (1U77) match closely those found in 1U7C and 1U7G, and were not included in the representation. (*C*) Key configurational deviations of the substrate binding site. The Zheng down-configurations of Q104 and F68 (*thin licorice*) are depicted along with the corresponding Khademi up-configuration of Q104. The S68 residue in the Khademi structures is not shown, instead the F68 up-configuration adopted in the starting structures for our initial simulations is depicted (see Table 4.1). (*D*) Assigned protonation states of H168, H318, and Amm adopted in our simulations. For sim-I the δ- and ϵ-nitrogens of H168 and H318 are labeled. The differently placed protons are colored orange. The snapshots are of energy-minimized systems from which our simulations were started. For easing visual comprehension, figures showing explicit residues in the protein structure are all viewed from the same angle as chosen for Fig. 4.1 *B*. Exceptions are Fig. 4.1 *C* and Fig. 4.3, *B* and *C*, which are rotated around the z axis.

binding site for NH_4^+/$MeNH_3^+$ at the periplasmic pore entrance [17, 18, 73].

The Khademi structure [Protein DataBank (PDB) entry No. 1U7G] is so far the only x-ray structure where Amm molecules are proposed to occupy the channel, i.e., one extraluminal NH_4^+ and three intraluminal gaseous NH_3 [17]. This is in accordance with experimental AmtB data, which consistently indicate-gradient driven NH_3 uniport, i.e., diffusion or passive transport [17, 29, 34]. However, controversy about which Amm species are recognized by AmtB prevails. The most common view is NH_4^+ recognition [17, 18, 34], but NH_3 recognition has also been proposed [29].

Noteworthy is that the NH_3 diffusion across AmtB is similar to that across bilayers. This questions the biological significance of AmtB albeit enhanced recruitment of Amm, i.e., NH_4^+, could be one justification. Amm transport mechanisms based on *E.coli* AmtB have been proposed [17, 18]. However, the detailed mechanism of NH_4^+ recruitment and NH_3 conduction, and their relationships to the structure of AmtB are unknown at the molecular level [18]. To gain such insight, we conducted classical mechanical molecular dynamics (MD) simulations of AmtB.

4.2 Materials and Methods

This section provides information about the simulation setup, chosen parameters, simulation details and pK_a calculations.

4.2.1 System Setup

The crystal structure of monomeric AmtB with the mutations F68S, S126P, and K255L, and all Methionines replaced by Selenomethionine (Mse) was obtained from the PDB as No. 1U7G [17]. We chose this crystal structure because Amm presumably occupies the substrate binding site and the channel lumen and because the two Zheng structures lacks coordinates for residues in loops connecting helix TM5 to TM6 and TM9 to TM10 [18].

Double side-chain coordinates (A and B) existed for residues Mse{200,328}, I15, S{43,155,335,358}, V{60,281,308}, L{88,259}, T203 and C326 [17]. We used the A-coordinates. All introduced mutations were backmutated to the appropriate wt residues. Due to the F68S mutation in the Khademi structures, Q104 takes a different configuration (up) than in the Zheng structures (down) as shown in Fig. 4.1C. All other residues defining the substrate binding site are identically positioned in all five *E.coli* AmtB crystal structures [17, 18]. For sim-{I–V} we adopted the up-configuration of Q104 from the Khademi structure 1U7G, which differs from the Zheng down-configuration by a rotation of the C_α-C_β-C_γ-C_δ dihedral. Due to steric hindrance of Q104 we placed the backmutated F68 side chain in an up-configuration above the neighboring H100 imidazole ring (Fig. 4.1C).

Conventional (pH=7) protonation states were chosen for all Aspartic and Glutamic acids and for Lysine residues. Based on visual inspection, N_ϵ protonation was chosen for all Histidine residues except for luminal H168 and H318. These two adjacent Histidines are positioned such that a proton is shared between their N_δ atoms. For classical mechanical MD simulations in explicit water the proton has to be permanently assigned to one of the two Histidines. We assigned protons to these Histidines in three different ways as shown in Fig. 4.1D (see also Table 4.1). For the charged state in sim-I, we determined to which N_δ atom of H168 and H318 the shared proton is most likely bound by estimating their pK_a values (see pK_a Calculations). Since protonation of H318:N_δ is favored over protonation of H168:N_δ, H318:N_δ was protonated in sim-I.

Hydrogen atoms were added to the monomer and crystal waters using PSFGEN distributed with NAMD [151]. The trimeric structure of AmtB was generated with VMD [152] using the transformation matrices provided in the PDB file. The symmetry related atoms 549, 678 and 767, all being oxygen atoms of water molecules, were not included in the transformation, but were added after the trimer was constructed. Cysteines C109 and C56 are close enough to form a disulfide bond, but the electron density map [17] does not show evidence of such a bond and it was therefore not introduced.

For the lipid membrane, POPE lipids (16:0/18:1c9) were used. This lipid resembles well the PE rich *E.coli* membrane [153]. In accordance with the experimental value for lamellar phase [154], a bilayer with 40 Å vertical separation between phosphorous headgroup atoms was constructed from 1920 *cis*-POPE lipids arranged in a regular hexagonal lattice with a lattice vector of 8.3 Å. The membrane plane was chosen parallel to the xy-plane. When positioning AmtB into the membrane, lipids overlaying or within 2 Å of the AmtB trimer (including crystal water) were deleted. An identical number of lipids were retained in the upper and lower leaflet. Crystal waters within the hydrophobic part of the bilayer (-15 Å $\leq z$ or $z \geq 15$ Å) but outside the protein interior were removed. SOLVATE [155] was subsequently used to add an ellipsoidal solvation shell around the membrane and the solvent-exposed parts of AmtB.

Given the trimeric structure of AmtB it is feasible to use a hexagonal simulation cell. A hexagonal periodic box was cut out in the x, y plane from the above system, while ensuring a minimum distance of 30 Å between the protein complex and the edges of the hexagonal box. The resulting hexagonal basis vectors (in Å) were $\mathbf{v}_1$=(141.1, 0.0, 0.0), $\mathbf{v}_2$=(70.6, 122.2, 0.0), and $\mathbf{v}_3$=(0.0, 0.0, 84.0). The corresponding dimensions of the AmtB trimer were 73.6 Å in the $\mathbf{v}_1$ and $\mathbf{v}_2$ directions, and 66.9 Å in the $\mathbf{v}_3$ direction. An appropriate number of water molecules were replaced by chloride counterions to bring the system to overall neutrality. The final system for sim-I consisted of an AmtB trimer (17.106 atoms), 9 NH_3 molecules (36 atoms), 3 NH_4^+ ions (15 atoms), 400 POPE lipids (50.000 atoms), 25.029

water molecules (75.087 atoms), and 12 Cl^- ions (12 atoms), adding up to a total system size of 142.256 atoms. When changing the protonation states of H168, H318, and NH_3 molecules (to NH_4^+), the number of counterions were adjusted accordingly.

4.2.2 Parameters

We used the CHARMM 27 parameter set [156]. The CHARMM neutral aliphatic amines parameters [157] were used for NH_3 molecules. No CHARMM parameters are available for NH_4^+ ions, but identical Lennard-Jones parameters are used for all nitrogen atom types in the CHARMM 27 parameter set, albeit different from those developed for NH_3 [157]. To be consistent with the existing force field, we decided to use the CHARMM 27 parameters of the ammonium nitrogen. Bond and bond-angle parameters were chosen to be the same as for the -NH_3^+ moiety of Lysine. For partial charges of NH_4^+, OPLS charges are available [158]. OPLS charges are as CHARMM charges optimized to reproduce experimental data. OPLS and CHARMM partial charges of -NH_3^+ in Lysine are nearly identical [156, 158]. Therefore we found it appropriate to use the OPLS partial charges for NH_4^+.

4.2.3 Simulation Details

Molecular dynamics simulations were conducted using NAMD [151]. Temperature and pressure were controlled by a Langevin thermostat and barostat, respectively. The integration time step was 1 fs and coordinates were saved every 0.5 ps. Electrostatic interactions were treated using the particle-mesh Ewald method, with a grid spacing <1.0 Å, using a multiple timestepping scheme. The van der Waals interactions were cutoff at 12 Å using a switching function starting at 10 Å. Full periodic boundary conditions were imposed.

Initially, the systems were minimized while keeping protein, Amm and crystal water fixed. An additional minimization without fixation was subsequently performed. The systems were then equilibrated for 250 ps while fixing the above-mentioned parts. After equilibration, the systems were again minimized without fixation. MD simulations without fixation were then performed for at least 2 ns at T=310 K and P=1 atm (NPT ensemble).

4.2.4 pK_a Calculations

When appropriate we carried out electrostatic pK_a calculations to estimate the pK_a shifts (ΔpK_a) of H168 and H318 in order to determine their preferred protonation states. In the calculations we included all protein and Amm partial charges. The partial charges were taken from CHARMM 27 and OPLS, respectively. Following the approach of Faraldo-Gomez and

Roux [159] we used a water probe of radius 1.4 Å to determine the protein solvent-accessible surface representing the dielectric boundary to the surrounding water and membrane. These were treated implicitly by means of dielectric constants of $80\epsilon_0$ and $2\epsilon_0$, respectively. The membrane was represented by a 40 Å thick slab centered at the geometrical center of the monomer. Due to overlap between the membrane slap and the two water filled vestibules leading into channel lumen, these were explicitly assigned a dielectric constant of $80\epsilon_0$. In all calculations we used a cubic grid with 1 Å grid spacing and with x, y, and z dimensions being at least twice the AmtB monomer dimensions. This ensured that an applied focusing procedure (same calculation, but with half the grid spacing, and boundary conditions given from the unfocused calculation) encapsulated the entire monomer. Before the calculations, any grid points not assigned above were assigned a dielectric constant of $80\epsilon_0$. The electrostatic potential was obtained by solving the Poisson-Boltzmann equation implemented in CHARMM [107]. From this $\Delta\mathrm{pK_a}$ values were derived.

For the intraluminal protonation state of sim-I, we had to determine to which N_δ atom of H168 and H318 in the starting structure the shared proton is most likely bound. For that purpose both Histidines were initially protonated at the N_ϵ position. pK_a shifts were then calculated for protonation at each N_δ and with different dielectric constants assigned to the protein ($\epsilon_p = \{2,4,6,8\}\epsilon_0$). In all cases the largest pK_a shift was obtained for protonation of H318:N_δ ($\Delta\mathrm{pK_a} = \{6.7,2.7,1.5,1.0\}$), leading to relative pK_a shifts [$\Delta\mathrm{pK_a}$(H318:N_δ)/$\Delta\mathrm{pK_a}$(H168:N_δ)] in the range 2.4-3.8.

To identify the preferred neutral protonation state of H168 and H318, we calculated ensemble-averaged $\langle\Delta\mathrm{pK_a}(\mathrm{H168{:}N}_\epsilon)\rangle$ and $\langle\Delta\mathrm{pK_a}(\mathrm{H318{:}N}_\epsilon)\rangle$ values for sim-I, sim-$\mathrm{II}^{\mathrm{wt}}_{\mathrm{noAmm}}$, and sim-$\mathrm{III}^{\mathrm{wt}}_{\mathrm{noAmm}}$. In none of the calculations did we include water explicitly. Thereby, $\Delta\mathrm{pK_a}$ values for a channel devoid of any substrate was obtained for sim-$\mathrm{II}^{\mathrm{wt}}_{\mathrm{noAmm}}$ and sim-$\mathrm{III}^{\mathrm{wt}}_{\mathrm{noAmm}}$. For all simulations in consideration $\langle\Delta\mathrm{pK_a}(\mathrm{H168{:}N}_\epsilon)\rangle$ and $\langle\Delta\mathrm{pK_a}(\mathrm{H318{:}N}_\epsilon)\rangle$ were calculated by changing the given intraluminal protonation state to that of sim-III and sim-II, respectively, and subsequently protonating either H168:N_ϵ or H318:N_ϵ. For these calculations a dielectric constant of $\epsilon_p = 2\epsilon_0$ was assigned to the protein. When no Amm is present in the lumen (sim-$\mathrm{II}^{\mathrm{wt}}_{\mathrm{noAmm}}$, sim-$\mathrm{III}^{\mathrm{wt}}_{\mathrm{noAmm}}$) $\langle\Delta\mathrm{pK_a}(\mathrm{H168{:}N}_\epsilon)\rangle$ is relatively larger than $\langle\Delta\mathrm{pK_a}(\mathrm{H318{:}N}_\epsilon)\rangle$, due to a stabilization of H168:N_ϵH by T273:O_γ. This is compensated when intraluminal Amm is present (sim-I), which then yields $\langle\Delta\mathrm{pK_a}(\mathrm{H318{:}N}_\epsilon)\rangle$ larger than $\langle\Delta\mathrm{pK_a}(\mathrm{H168{:}N}_\epsilon)\rangle$.

4.3 Results and Discussion

In our simulations we initially considered five different intraluminal protonation states as shown in Fig. 4.1D and as summarized in Table 4.1. We use the term 'intraluminal protonation state' for the combined protonation states of H168, H318, and intraluminal Amm. We correspondingly refer to these simulations as sim-I, sim-II, sim-III, sim-IV, and sim-V, or sim-{I–V}. In all five simulations one NH_4^+ is positioned extraluminally at the substrate binding site. In sim-I a positive charge is localized at H318. The simulations sim-II and sim-III are both of neutral tautomeric H168 and H318 protonation states. The simulations sim-IV and sim-V correspond to sim-II and sim-III but with one intraluminal NH_3 substituted by NH_4^+. We also conducted simulations of intraluminal protonation states identical to sim-II and sim-III, but with changes in the protein structure at the periplasmic NH_4^+ binding site (superscript *wt*), with protonated H100 (subscript $H100^+$), and with the channel devoid of Amm (subscript *noAmm*).

In the following we discuss *1*) recruitment and binding of NH_4^+ to the substrate binding site, *2*) NH_4^+ transfer from the substrate binding site to pore lumen, *3*) intraluminal Amm dynamics, *4*) Amm exit to cytoplasm, and *5*) channel hydration. Finally we propose a plausible mechanism integrating NH_4^+ recruitment and translocation to pore lumen, and the backward relay of the intraluminally recruited proton to the periplasm leading to net transport of NH_3 only.

Simulation	**Lumen**		**Substrate**				**Substrate binding site**			**Amm**	**Length**
	[H168:N_ϵ	H318:N_ϵ]	[Am1	Am2	Am3	Am4]	[F68	Q104	H100:N_ϵ]	**constrained**	**(ns)**
sim-I	H	H (+)	NH_4^+	NH_3	NH_3	NH_3	up	up	H	no	2.2
sim-II	H		NH_4^+	NH_3	NH_3	NH_3	up	up	H	no	1.9
sim-III		H	NH_4^+	NH_3	NH_3	NH_3	up	up	H	no	3.0
sim-IV	H		NH_4^+	NH_3	NH_3	NH_4^+	up	up	H	no	2.0
sim-V		H	NH_4^+	NH_4^+	NH_3	NH_3	up	up	H	no	1.9
$\text{sim-III}^{\text{wt}}_{\text{fix}}$		H	NH_4^+	NH_3	NH_3	NH_3	down	down	H	yes	2.0
$\text{sim-III}^{\text{wt}}$		H	NH_4^+	NH_3	NH_3	NH_3	down	down	H	no*	4.0
$\text{sim-III}^{\text{wt}}_{\text{H100}^+}$		H	NH_4^+	NH_3	NH_3	NH_3	down	down	H (+)	no*	4.0
$\text{sim-II}^{\text{wt}}_{\text{noAmm}}$	H		-	-	-	-	down	down	H	-	10.0
$\text{sim-III}^{\text{wt}}_{\text{noAmm}}$		H	-	-	-	-	down	down	H	-	10.0

Table 4.1: Summary of the simulations. The lumen, substrate, and substrate binding site configuration of each simulation and applied constraints and simulation lengths. H indicates N_ϵ-protonation of Histidines H100, H168, and H318, while blank (no H) indicates N_δ-protonation. The +-symbol indicates positively charged Histidines with both imidazole nitrogens protonated. Am1–Am4 are substrate locations in the Khademi structure 1U7G [17]. The terms "up" and "down" denote distorted and wt configurations, respectively, of the substrate binding site residues F68 and Q104 (see text). "Constrained" implies that Amm molecules are fixed at the substrate locations in the x-ray structure.
*Restarted from $\text{sim-III}^{\text{wt}}_{\text{fix}}$ with Amm released.

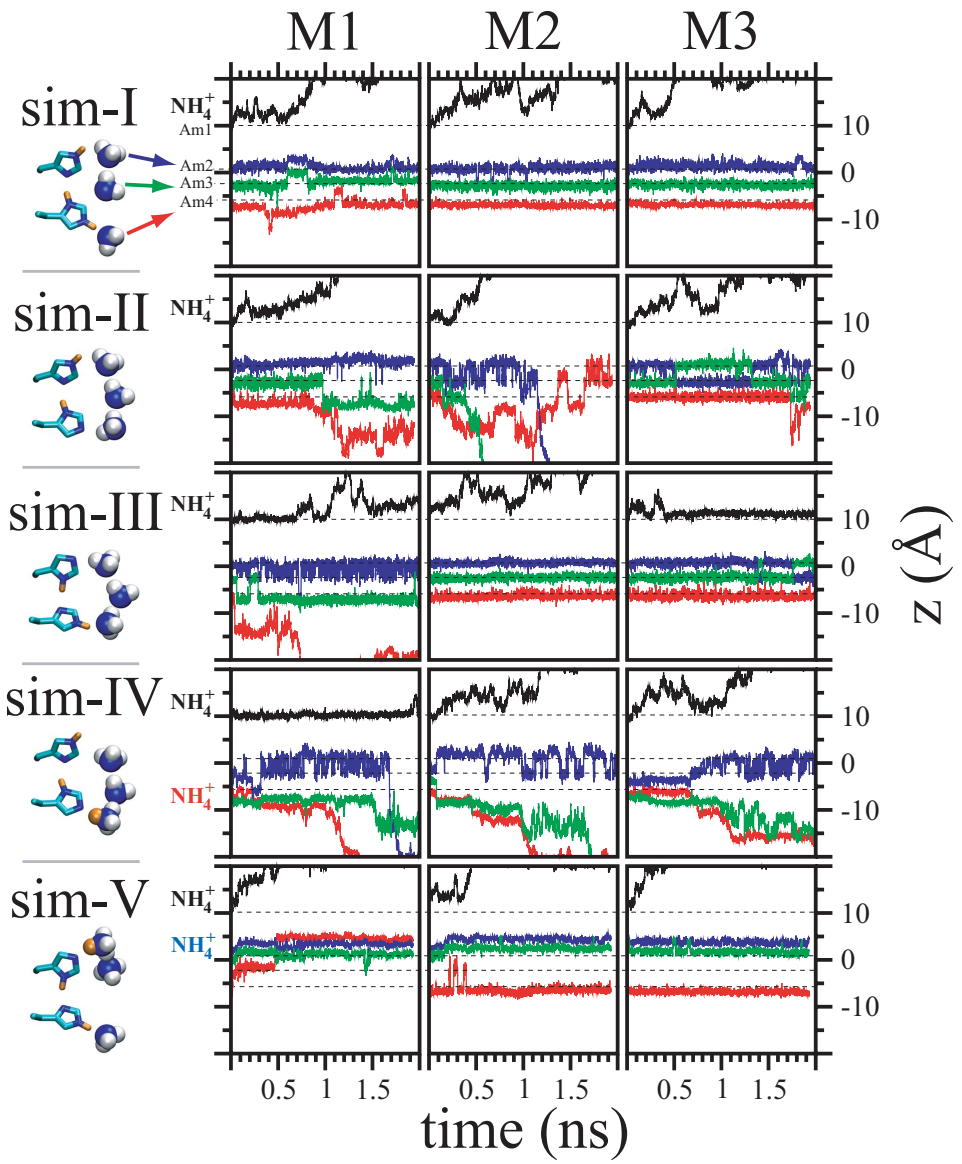

Figure 4.2: Trajectories. Amm z-coordinates as function of time for simulations sim-{I–V}. (*Black*) NH_4^+ initially at the Am1 position. (*Blue*) NH_3/NH_4^+ initially at the Am2 position. (*Green*) NH_3 initially at the Am3 position. (*Red*) NH_3/NH_4^+ initially at the Am4 position. M1, M2, and M3 refer to the individual monomers in the AmtB trimer. The Am1–Am4 positions are indicated by dashed lines. The strongest correlation between Amm molecules is observed between NH_4^+ and NH_3, e.g., red and green curves in sim-IV, and blue and green/red curves in sim-V. Occasionally, some correlation is observed between the two NH_3 closest to the periplasm. These may however also interchange positions; see, e.g., blue and green curves of monomer M3 in sim-II and sim-III.

4.3.1 Interactions at the Substrate Binding Site

Trajectories (z-coordinates) of the extraluminal NH_4^+ and of the three intraluminal Amm are shown for sim-{I–V} in Fig. 4.2. Surprisingly, the NH_4^+ ion initially located at the Am1 position and partially coordinated to water was generally found to leave this position. To elucidate if this observation was biased by our starting conditions, we initially constrained NH_4^+ to the Am1 position for 2 ns, which, however, led to the same result (sim-III_{fix}^{wt}, Table 4.1, data not shown). The escape of NH_4^+ from the substrate binding site could be due to shortcomings of the used force field, i.e., a possible poor description of cation-π interactions between NH_4^+ and aromatic residues at the substrate binding site [160]. These are only implicitly accounted for by the nonbonded interaction terms in the CHARMM energy function. Cation-π interactions can be treated explicitly by using polarizable force fields or *ab initio* methods. We have conducted *ab initio* calculations of the substrate binding site based on structures generated from *ab initio* MD (Carr-Parrinello MD) simulations of AmtB. Our results indicate that cation-π interactions are not the main contribution to the binding of NH_4^+ (data not shown here, see Chapter 6). Thus, we do not think that the easy escape of NH_4^+ should be assigned as a shortcoming of the force field, rather it should be interpreted as an effect of the substrate binding site being a low-affinity cation binding site. When close to the Am1 position, we observe that NH_4^+ is involved in one or more prominent hydrogen bond (H-bond) interactions with F161:O, A162:O, W148:N_ϵ, D160:O_δ, D160:O, S219:O, and S219:O_γ, as shown in the histogram of NH_4^+-protein interactions in Fig. 4.3A. These interactions reflect NH_4^+-protein interactions at positions slightly displaced from the Am1 position as shown in Fig. 4.3B. Most interesting is the interaction with A162:O, which we will refer to as A162:C=O or as A162-G163 when referring to the entire peptide bond. In the crystal structures A162:C=O points in-between the π-stacking side chains of F107 and F215 (Fig. 4.3C, *inset*). In our simulations A162:C=O reorients to point either into the substrate binding site (up) or into the pore lumen (down) (Fig. 4.3C). The role of A162-G163 is further discussed below. Also interesting is the interaction between NH_4^+ and D160. Upon rotation of the carboxylate group, D160 becomes a hydrogen acceptor to NH_4^+ (Fig. 4.3B), while it still retains one of its two H-bonds that fix the irregular helical structure at the periplasmic constriction region (Fig. 4.1B). The mutation D160A inactivates AmtB, suggesting that D160 is involved in Amm binding and/or transport [59]. The role of D160 has also been proposed to be structural in fixing surrounding residues [17, 18]. Our simulations support both views, i.e., that D160 participates in ordering the surrounding protein structure, and, in recruiting and binding of Amm; first via its carbonyl oxygen and subsequently via its carboxylate group.

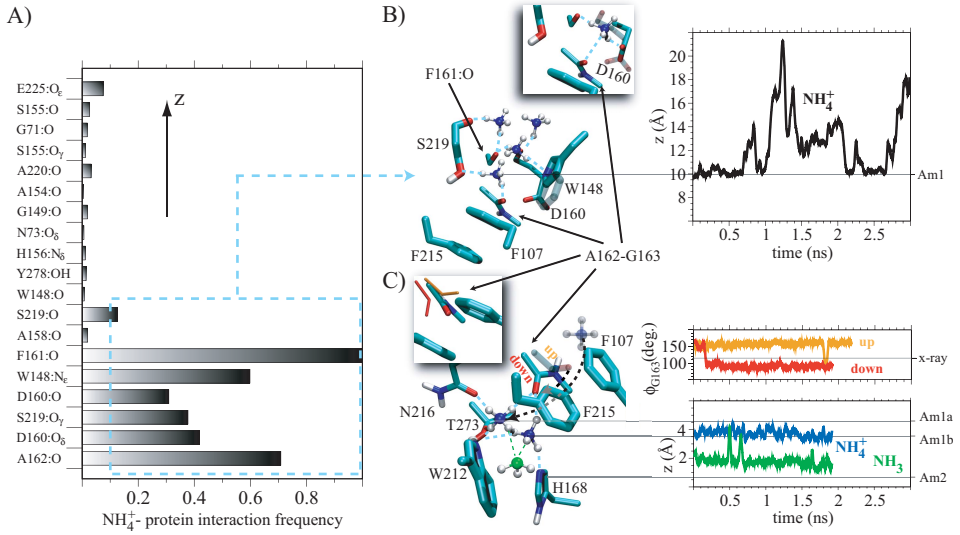

Figure 4.3: NH_4^+ recruitment and translocation. (*A*) Histogram of H-bonding interactions in sim-{I–V} between NH_4^+ and residues in the periplasmic vestibule. The residues are ordered according to their z-coordinates. The histogram was constructed by counting the occurrences of NH_4^+ within protein distances below 3 Å and normalizing the maximum count to 1; (*B*) Five snapshots of NH_4^+ (un-)binding at the substrate binding site along the sim-III trajectory. The snapshots are of representative NH_4^+ positions, which cause the most prominent H-bond interactions of the histogram in *A*. The inset shows the binding state where the D160 carboxylate group is rotated to accept a NH_4^+ hydrogen. The right panel shows the trajectory (z-coordinate) of NH_4^+ in monomer M1. The snapshots were aligned by minimizing the RMSD between all C_α atoms; (*C*) Two snapshots of NH_4^+ at intraluminal binding sites in the periplasmic constriction region (sim-V). The additional NH_4^+ (*transparent*) corresponds to the bottom position in *B* with A162:C=O in the up-conformation (*transparent*). In the up-configuration A162:C=O can form a H-bond to water if no NH_4^+ is present (not shown). When NH_4^+ is positioned intraluminally A162:C=O reorients to a down-conformation (*solid*). We name the two intraluminal NH_4^+ positions Am1a and Am1b. In the Am1a state NH_4^+ H-bonds to A162:C=O, W212:N_ϵ/T273:O_γ, N216:O_δ, and NH_3. In the Am1b state NH_4^+ H-bonds to A162:C=O, H168:N_ϵ, W212:N_ϵ/T273:O_γ, and NH_3. The NH_4^+ hydrogen coordinating to W212:N_ϵ/T273:O_γ shifts between these two hydrogen acceptors, which are themselves H-bonded via W212:N_ϵH. The black dashed arrow indicates the direction of NH_4^+ translocation across the F107/F215 stack. An opening of the stack can be recognized. The inset shows the x-ray structure position of A162:C=O and the corresponding closed conformation of the F107/F215 stack. The right panels show ϕ_{G163} trajectories of sim-III, monomer M1 (*orange* and mainly "up") and sim-V, monomer M3 (*red* and mainly "down"), and the z-coordinates of NH_4^+ (*blue*) and NH_3 (*green*) in sim-V, monomer M3. Note that the viewing angle in *B* and *C* is different from Fig. 4.1 *B*.

4.3.2 Q104 Influence on Substrate Binding Site Occupation

Simulations sim-{I–V} generally show rapid escape of extraluminal NH_4^+ (Fig. 4.2) and slow relaxation of F68 and Q104 toward their wt down-configurations (Fig. 4.1C). It is therefore of interest to examine whether or not F68 and Q104 influence substrate recruitment and occupation of the substrate binding site, since the functional importance of this recruitment site has remained unclear [17, 18, 73]. Accordingly, we changed the configurations of F68 and Q104 in the Khademi structure 1U7G to the down-configurations as observed in the Zheng structures. We used the same intraluminal protonation state as in sim-III and simulated initially with Amm fixed for 2 ns (sim-III^{wt}_{fix}) and subsequently continued for 4 ns with Amm released and H100 being either neutral (sim-III^{wt}) or protonated (sim-$III^{wt}_{H100^+}$). The solvent-accessible H100, fixed by a hydrogen bond to Y64:H_γ, may be protonated given the pH in *E.coli* periplasm [$pH_{periplasm} < pH_{cytosol} \simeq 7.6$ [161]] and thereby affect the dynamics of neighboring F68 and Q104. However, regardless of the protonation state of H100, none of the above simulations revealed any significant correlation between NH_4^+ occupation of the substrate binding site and the F68/Q104 dynamics.

The F68/Q104 dynamics might be important for recruiting substrates other than NH_4^+, and was therefore further investigated. We find that Q104 adopts several different up-configurations, which are not related to specific H-bond interactions, neither with the substrate nor with the protein, e.g., with H100. These observations in wt AmtB simulations dispute that Q104 up-configurations are solely caused by the F68S mutation [18]. Rather, we find that they also may result from steric effects, e.g., the space made available by F68 and H100. Since we do not observe Q104 to affect binding of NH_4^+ we propose that Q104 by means of sterical effects influence the binding of the methyl group in MeAmm. This may explain why Amm is a better AmtB substrate than MeAmm [29, 34].

4.3.3 Entry and Exit of NH_4^+ at the Substrate Binding Site

When NH_4^+ leaves the substrate binding site we find that it becomes fully hydrated (Fig. 4.5A) and diffusively migrates out of the periplasmic vestibule within 0.5–1.5 ns (Fig. 4.2). The calculated diffusion constant is 0.83×10^{-5} cm^2/s, which is equal to that of free diffusion $0.82(\pm0.02)\times10^{-5}$ cm^2/s [162]. While NH_4^+-protein interactions are specific at the substrate binding site, they are less specific in the periplasmic vestibule. There, we find that NH_4^+ mainly interacts with backbone carbonyl groups (Fig. 4.3A–B). Further out the vestibule a somewhat more specific interaction is again observed; namely with the carboxylate group of E225, which, located within one of four loops surrounding the periplasmic vestibule, protrudes into the

solvent. The other three loops have residues E70, E87, and E344 located in similar positions (Fig. 4.4A), and these four Glutamic acids, which we modeled in their charged form, may therefore be considered as NH_4^+ capturing residues.

4.3.4 A162 Guided Ammonium Translocation

Pore blocking F107 and F215 have identical π-stacking conformations in all five crystal structures [17, 18]. Another common feature is that the backbone carbonyl group of the conserved A162 is directed toward the space in-between the F107/F215 stack (Fig. 4.3C, *inset*).

In our simulations A162:C=O predominantly takes a conformation pointing up toward and into the substrate binding site, forming a H-bond to either extraluminal NH_4^+ or water (Fig. 4.3C). A striking exception from this observation is sim-V where we introduced intraluminal NH_4^+, which occupies the intermediate positions Am1a–b between Am1 and Am2 (Fig. 4.3C). In this case A162:C=O shifts to a down-conformation, thereby forming a H-bond to intraluminal NH_4^+, which in turn is stabilized significantly at this position (Fig. 4.3C). A162:C=O moves 3–4 Å along z when shifting between up- and down-conformations. As depicted in Fig. 4.3C, the shift is facilitated by the adjacent G163 via a change of its ϕ-torsion from $\sim$160° to $\sim$95°. A162:C=O has been speculated to be a hydrogen acceptor for substrate molecules in a transient open state of the constriction region [18]. Based on our two observed conformations we suggest that the function of A162:C=O is to actively guide NH_4^+ across the periplasmic constriction, i.e., across the F107/F215 stack via direct H-bonding interactions.

Intraluminal NH_4^+ in the upper part of the pore lumen (sim-V) is strongly stabilized by tetrahedral coordination to channel-lining residues. Three of the NH_4^+ hydrogens are coordinated to A162:C=O, W212:N_ϵ/T273:O_γ, and NH_3, respectively, while the remaining hydrogen coordinates either to N216:O_δ or to H168:N_ϵ (Fig. 4.3C). Given this intraluminal stabilization, in particular at position Am1b, we suggest that NH_4^+ may translocate from the substrate binding site to the Am1b position before deprotonation, rather than to the Am2 position as previously proposed [17]. Both scenarios require that H168:N_ϵ is deprotonated, implying that neutral H168:N_δH-H318:N_ϵH (sim-{III,V}, Fig. 4.1D) is the native protonation state of the intraluminal Histidines before any proton relay involving these residues.

In all our simulations, correlated side-chain rotations of 180° of F107 and F215 are frequently observed, except in sim-V, where the side chains are rotated and rotationally constrained. This results in a more open state of the constriction region indicating that translocation of NH_4^+ from the substrate binding site to the pore lumen is coupled to F107 and F215 side chain rotations, which supports previous suggestions [17, 18].

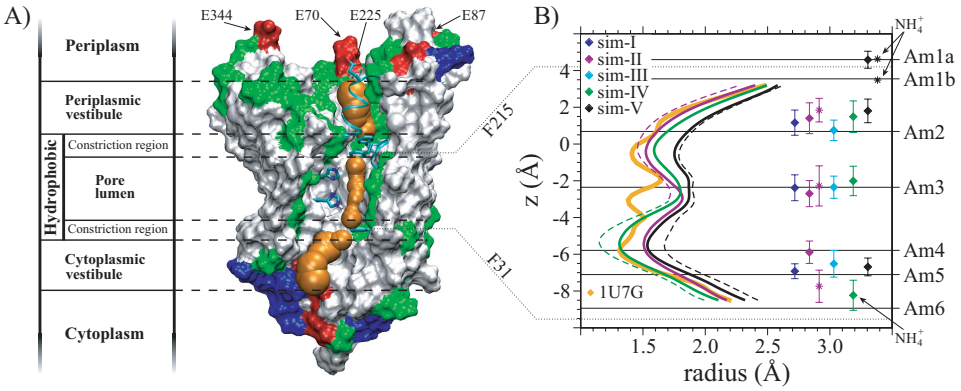

Figure 4.4: Channel characteristics and intraluminal Amm positions. (*A*) Side view of AmtB and definition of channel regions [17, 18]. The protein is shown in surface representation and colored according to residue types; red is acidic, blue is basic, green is polar, and white is nonpolar. Four Glutamic acids, which surround the periplasmic vestibule and might be important in attracting NH_4^+ to the substrate binding site, are pointed out. Orange spheres represent the pore with their radii calculated using HOLE [19]; (*B*) The intraluminal pore radii for the Khademi 1U7G structure (*orange*), and the average pore radii for sim-{I–V} over the last 1 ns of each simulation (also averaged over monomers) are shown. Dashed lines are maximal upper (sim-V) and lower (sim-II and sim-IV) limits of the radii calculated from the respective standard deviations. Average positions of intraluminal Amm are indicated by diamonds and stars. For each of sim-I, sim-III, and sim-IV, only one entry (*diamonds*) is shown representing averages over all three monomers. For sim-II, the diamonds correspond to monomer M3, where all NH_3 are present in the channel lumen, while stars represent an average of M1 and M2 data, where the channel lumen is partially occupied. For sim-{IV,V} intraluminal NH_4^+ positions are pointed out.

4.3.5 The Hydrophobic Pore and the L114 Cavity

Fig. 4.4A shows a vertical cut through the Khademi structure 1U7G and the channel pathway calculated using HOLE [19]. The intraluminal channel radii for the Khademi structure and sim-{II,IV,V} are shown in Fig. 4.4B. The three latter curves represent the smallest (sim-{II,IV}) and largest (sim-V) radii of all conducted simulations. In general, the channel widens relative to the crystal structure. The largest difference is observed around the Am3 position and is due to a conformational change of the L114 side chain, which creates a cavity adjacent to that position. This observation is a general feature, since that region exhibits small deviations in radii when comparing our simulations in Fig. 4.4B. Larger deviations in radii are observed around the Am2 and Am4 positions. These are due to nonspecific small displacements of M23, V67, I110, H168, W212, and F215, and F31, I28, L114, L208, and I266, respectively. The largest deviations in radii between simulations do, however, not exceed 0.6 Å, which we consider as minor channel radii fluctuations.

4.3.6 A Native Protonation State of the Intraluminal Histidines

In general, the intraluminal Amm positions are mutually correlated. The strongest correlation is observed between NH_4^+ and NH_3, but occasionally also between the two NH_3 molecules closest to the periplasm (Fig. 4.2). These two NH_3 can interchange positions. Interchange may occur when both NH_3 molecules simultaneously occupy the Am3 position and the adjacent L114 cavity, where NH_3 is stabilized by the hydrogen acceptor I110:O. The NH_3 distributions center around the Am3 position (Fig. 4.4B) and exhibit no significant dependence on the protonation state of the intraluminal Histidines nor on the occupation of the Am2 and/or Am4 positions. This does not apply to NH_3 positioned at the periplasmic or cytoplasmic boundaries, respectively (Fig. 4.4B).

In sim-IV and sim-V the introduced intraluminal NH_4^+ never occupies the middle of the pore (Fig. 4.2). This is in accordance with the suggestions of NH_3 being the conducted species and with the barrier against cation conduction being highest in this region [17, 18, 34].

In the Khademi structure 1U7G, three NH_3 molecules occupy the Am2, Am3, and Am4 positions (Fig. 4.1B). It is argued that H168:N_ϵH is hydrogen-donor to T273:O_γ, while H318:N_ϵ is hydrogen-acceptor to Amm positioned at Am4 [17], which corresponds to the intraluminal protonation state of our sim-II (Fig. 4.1D). However, in this Khademi structure, a β-octylglucoside blocks the cytoplasmic vestibule all the way up to F31, and the aromatic side chain of F31 is slightly lifted upwards relative to the Zheng structures [17, 18]. When comparing the average positions of intraluminal NH_3 in our simulations, we find that sim-III is in best accordance with the Amm positions found in the Khademi structure 1U7G (Fig. 4.4B). In sim-III, the most cytoplasmic NH_3 is centered between Am4 and Am5, and conforms to both positions when considering the standard deviations. However, that minor inconsistency with respect to the crystal structure can be explained by the absence of β-octylglucoside in our simulations. Seemingly, presence of the β-octylglucoside shifts the position of the most cytoplasmic NH_3 upward to the Am4 position. Thereby the Khademi structure 1U7G conforms to the idea that the intraluminal protonation state of sim-III is the native state before proton relay. This is supported by our pK_a calculations (see Materials and Methods for details), which yield a larger shift in pK_a (ΔpK_a) for H318:N_ϵ relative to that for H168:N_ϵ, when assuming NH_3 (or water) molecules are occupying the pore lumen. In our simulations, NH_3 and/or water molecules occupy the pore lumen. Thus, we conclude that neutral H168:N_δH-H318:N_ϵH is the native protonation state of H168 and H318 in AmtB.

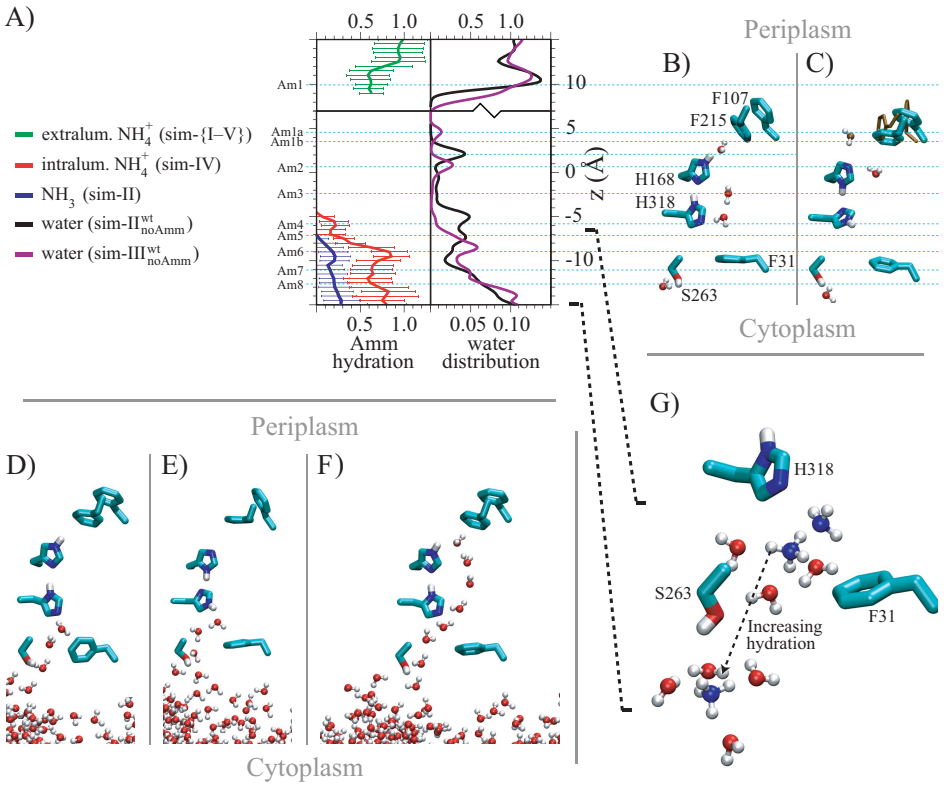

Figure 4.5: Channel hydration. (*A*) The fractional Amm hydration in sim-{I–V} and the distribution of intra- and extraluminal water in sim-II^{wt}_{noAmm} and sim-III^{wt}_{noAmm}. The fractional hydration of Amm was obtained by counting the number of water molecules within 2 Å of Amm during simulations, averaging this number for bins of size 0.5 Å along the z direction, and normalizing to the average bulk hydration. Error bars represent the fractional Amm hydration ± SD. The water distribution was obtained by counting the number of water molecules within cylindrical slabs of height 0.1 Å along the z direction, and then normalizing by setting bulk water slabs to 1. Due to low water densities in intraluminal slabs the lower graph uses a different scale for the water distribution. Values exceeding 1 arise due to immobilization of water in the periplasmic vestibule (above the substrate binding site Am1) in combination with a small size of the cylindrical slab volume used in our analysis. (*B*) Three water molecules accommodate the pore lumen simultaneously for 1.2 ns in sim-II^{wt}_{noAmm}; (*C*) Maximally, water protrudes up to the Am1a and Am1b positions (brown), but cannot stabilize an open state of the F107/F215 stack. Mainly water accommodates the Am2 position, sim-III^{wt}_{noAmm}; (*D*,*E*) Water enters into the pore lumen via single file formation to either H318:N_ϵ or H318:N_ϵH. The single file is stabilized by S263:O_γ(H); (*F*) A single file of water can extend all the way up to the periplasmic constriction region, sim-II^{wt}_{noAmm}; (*G*) NH_4^+ hydration and exit from the pore lumen in sim-IV. NH_4^+ forms a pair with NH_3 and is coordinated to three water molecules. When translocating toward the cytoplasm NH_4^+ becomes fully coordinated to water and so does NH_3 (not shown). Occasionally, NH_4^+ escape promotes NH_3 escape as well.

4.3.7 Hydration and Escape of Amm

Fig. 4.5A shows the average fractional hydration numbers of NH_3 (sim-II) and NH_4^+ (sim-IV) upon escape from the channel lumen. NH_4^+ hydration at an intraluminal position is exceeding that of a correspondingly positioned NH_3. In hydrated form, NH_4^+ leaves the lumen and escapes to the cytoplasmic vestibule (Fig. 4.5G). Occasionally, a NH_4^+-NH_3 pair is formed (Fig. 4.5G) and NH_4^+ escape thereby promotes NH_3 escape as well. The first specific interaction observed between NH_4^+ and the cytoplasmic vestibule is to S263:O_γ. Further out the vestibule, specific interactions with the D313 and D309 carboxylate groups also occur. It has been suggested that NH_3 may accept a proton close to the position of the D313 carboxylate group [17]. Of note is that this suggestion was based on the 1U7G structure where the cytoplasmic vestibule is blocked by β-octylglucoside. As discussed below, we observe in our simulations that the cytoplasmic vestibule is hydrated up to S263 (Fig. 4.5B–F). The crystal structures without β-octylglucoside also have water molecules at the Am5–Am8 positions (Fig. 4.1B), which are closer to the cytoplasmic constriction than D313 [17, 18]. This suggests that NH_3 is more likely to acquire a proton before reaching the D313 carboxylate group region, probably already when positioned at the Am4–Am5 positions.

We do not observe that Amm escape is correlated with large conformational changes in the N-terminus of TM10 or large V314 displacements as suggested previously [18]. Although we cannot rule out long-time protein conformational changes, this seemingly implies that the so-called "open" Zheng conformation (1XQE) might not be functionally relevant since the "closed" conformation, adopted by all other crystal structures, does not prevent partial Amm conduction according to our results.

4.3.8 Channel Hydration and Single File Water

To investigate channel hydration, we conducted two 10 ns simulations; sim-II^{wt}_{noAmm} and sim-III^{wt}_{noAmm}, with the channel devoid of Amm. The water distribution in Fig. 4.5A shows that water is present just above the Am1 position and may protrude 2–3 Å below, confirming that the Am1 position is accessible to solvent [17]. Fig. 4.5A also shows that in both simulations water enters the channel lumen, which confirms the suggestion of a partial channel hydration [18], but is at variance with related simulation results [17]. We observe that up to three water molecules can occupy the lumen simultaneously and access all positions (Am2–Am5) therein (Fig. 4.5B–C). However, intraluminal water molecules are predominantly positioned at the top and bottom of the channel lumen. The average water positions (Fig. 4.5A) depend in turn on the protonation state of the two luminal Histidines as exemplified in Fig. 4.5B–C. Interestingly, in both simulations, water entrance occurs by temporary formation of single file water extending from the cyto-

plasmic vestibule up to H318:N_ϵ and H318:N_ϵH, respectively (Fig. 4.5D–E). Moreover, a single file of water was also observed to form from the cytoplasm and all the way up to the periplasmic constriction region (Fig. 4.5F). In sim-{I–V} with Amm present in the channel, we observe water entry into the lumen also solely from cytoplasm. In all cases, S263:O_γ promotes formation of single file water inside the channel (Fig. 4.5D–F).

Single file water can conduct protons [163]. Even if water mediated proton conduction through the hydrophobic environment in the middle of the channel lumen is energetically unfavorable [18], at least proton conduction between cytoplasm and H318:N_ϵ is possible. Thus, proton conduction by single file water provides a possible mechanism for H^+ exchange with H168/H318.

No water crossing of the periplasmic hydrophobic constriction was observed in any simulation in accordance with the fact that AmtB does not sustain water transport [17]. Experiments also show that the channel does not conduct cations other than Amm and MeAmm. Water and cation filtering mechanisms are discussed below.

4.3.9 Mechanisms

To prevent leakage of water and cations, a two-filter system presumably has to be present in AmtB; one filter that excludes water, and one that excludes cations other than NH_4^+ or $MeNH_3^+$ from passing through.

Water Filtering at the F107/F215 Stack.

Our simulations indicate that specifically the periplasmic hydrophobic constriction is impermeable to water. Transfer of any substrate across the constriction region presuppose rotation of F107 and F215 (Figs. 4.3C and 4.5C). We observe that NH_4^+ can induce and stabilize a transient open state of the F107/F215 stack for several nanoseconds (Fig. 4.3C), while neither NH_3 nor water can (Fig. 4.5C). Thus, we propose that the F107/F215 stack and the otherwise hydrophobic character of the constriction region constitutes a filter against water transport.

Cation Filtering at Intraluminal NH_4^+ Binding Sites

A water filter should also filter NH_3; a hydrogen donor/acceptor and polar molecule like water. Combined with the observed stabilization of the F107/F215 stack mentioned above, this suggests that NH_4^+ is the Amm species transported across the constriction region. This is supported by free energy calculations, which identified cationic binding sites on both the extra- and intraluminal sides [18]. A cation filtering function of the constriction region was also suggested, but only selectivity against Na^+ ions, while not against K^+ ions, could be explained [18]. We have identified two

intraluminal binding sites both providing tetrahedral coordination of NH_4^+. In other ion channels, binding sites selective of Na^+ require 5–6 coordinating oxygens, while those selective of K^+ require eight coordinating oxygens and a larger binding cavity due to its larger size [164, and references therein]. Since NH_4^+ is roughly the same size as K^+ one may speculate that AmtB, due to the size of its intraluminal binding cavities (i.e., it accommodates both NH_4^+ and the larger $MeNH_3^+$), cannot coordinate Na^+ efficiently, as observed for other selectivity filters [164, and references therein]. The coordination provided by AmtB probably cannot stabilize K^+ either. Thus, we speculate that the tetrahedral coordination in AmtB is selective of NH_4^+, i.e., it counterbalances the dehydration cost of NH_4^+ better than that of other cations. In conjunction with the hydrophobic pore lumen, which prevents cation transfer [17, 18], this provides an intriguing cation filter mechanism. We do not hereby exclude that the F107/F215 stack may also provide some discrimination among cations.

Luminal Ammonium Binding and Proton Transfer

Since the recruited Amm species is NH_4^+, while the transported Amm species is NH_3, proton transfer reactions must occur at the channel entrance. According to our discussion above, deprotonation of NH_4^+ takes place after NH_4^+ transfer across the F107/F215 stack and prior to further translocation down the pore. As previously argued, the protonation state of luminal Histidines before proton relay is assumed to be the neutral H168:N_δH-H318:N_ϵH state. Since AmtB has been shown not to conduct protons [17], a proton must be transferred back to the periplasm. This proton transfer can be mediated by A162:C=O as shown in Fig. 4.6, which in our simulations forms a conserved H-bond with NH_4^+ when it is bound in the periplasmic constriction region. First A162:C=O guides NH_4^+ across the F107/F215 stack (Fig. 4.6, step *1*). When NH_4^+ occupies the Am1a or Am1b positions, we find that a perfectly aligned H-bond can be established between the backbone NH of the A162-G163 peptide bond and D160:O_δ (Fig. 4.6, step *2*). In general, hydrogen exchange of NH can be catalyzed by a base, i.e., deprotonation of NH, or by an acid, i.e., protonation of either O or N. The O protonation; the so-called imidic acid mechanism, is preferred in proteins [165, 166]. In our case, both intraluminal NH_4^+ and partially solvent accessible D160:O_δ may be considered as a stronger acid and a stronger base, respectively, when compared to their solvated counter parts. Thus, they can catalyze the hydrogen exchange via the imidic acid mechanism as a combination of base and acid catalysis (Fig. 4.6, step *2*). A similar mechanism has been proposed to be involved in proton pumping in cytochrome *c* oxidase [167, 168]. When the proton is effectively transferred from NH_4^+ to D160:O_δ it yields an intermediate state where all involved parts are neutral and A162:C-O-H reorients to the substrate binding site (Fig. 4.6, step *3*). The reorientation and thereby a less

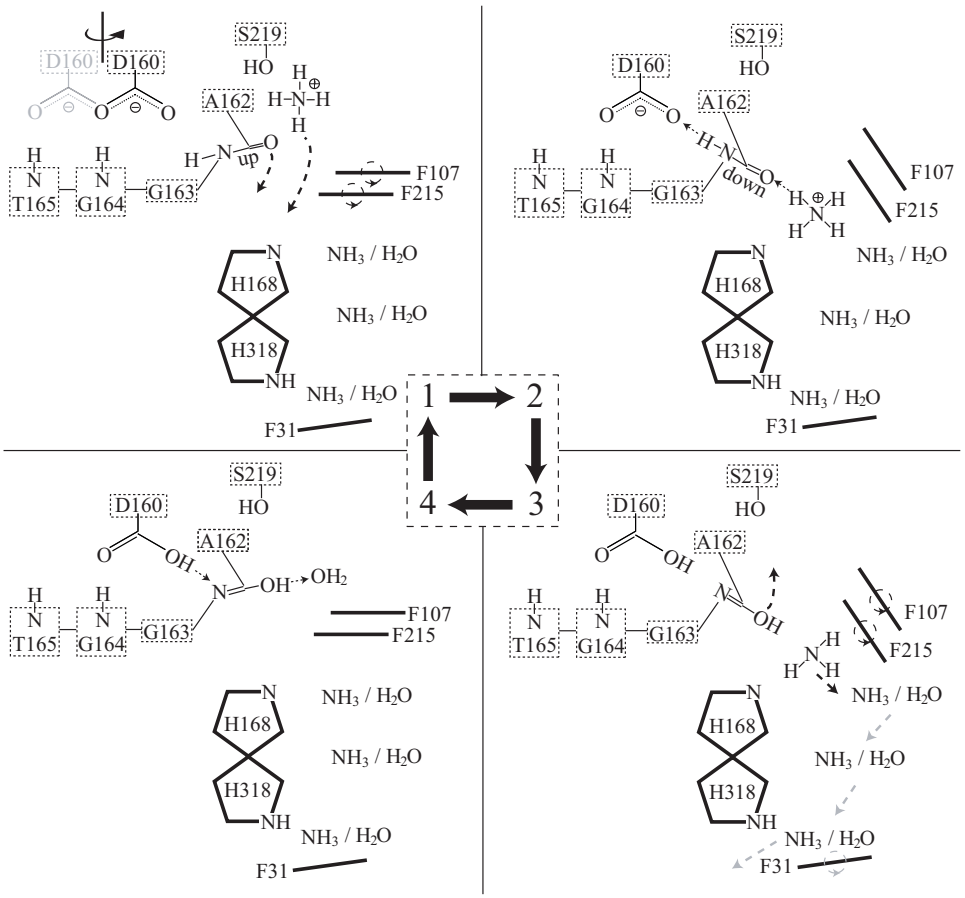

Figure 4.6: Proton transfer mechanism. (*1*) At the substrate binding site NH_4^+ H-bonds to S219:O_γ and A162:C=O. D160 can interact with NH_4^+ and G163:NH by rotation of its carboxylate group. In correlation with F107 and F215 rotation, NH_4^+ translocates to the channel lumen guided by A162:C=O; (*2*) When A162:C=O points into the lumen, G163:NH aligns to form a H-bond with D160:O_δ. Hydrogen exchange at the A162-G163 peptide bond is then catalyzed by D160 and NH_4^+, resulting in a net proton transfer from NH_4^+ to D160:O_δ, thereby yielding an imidic acid intermediate; (*3*) Charges on D160 and NH_4^+ are neutralized. NH_3 translocates further down the lumen. At some point the A162-G163 peptide bond reorients to the substrate binding site. Translocation of NH_3 may occur in any step and the gray arrows merely indicate one of several possible translocation mechanisms; (*4*) At an intermediate step between step *3* and step *1* the imidic acid mechanism is reversed due to a less favorable H-bond between D160:O_δH and G163:N, and a proton is transfered to water thereby recovering the initial state.

favorable interaction between G163:N and D160:O_δ-H, due to a changed direction of the H-bond, both promote proton transfer to a periplasmic water molecule to recover the initial state (Fig. 4.6, step *4*). This proposed proton transfer mechanism also adds a new explanation to the experiment where Amm conduction is lost upon D160 mutation; e.g. D160A; only the D160E mutation retains function [59]. Additionally, we have observed that rotation of the D160 carboxylate group may occur when NH_4^+ is bound at the substrate binding site. This provides an appealing coupling between NH_4^+ recruitment and the proton transfer mechanism described above. Our simulations do not provide information about the translocation mechanism of Amm further down the pore lumen. In Fig. 4.6 (step *3*) only one of several possible translocation mechanisms of intraluminal NH_3 is indicated. However, a mechanism that couples NH_4^+ entrance to the escape/protonation of an Am4–Am5 positioned NH_3, via the H168-H318 ring system, is also alluring.

4.4 Conclusion

We have conducted molecular dynamics simulations of *E.coli* AmtB with different protonation states of intraluminal H168, H318, and Amm. Additional simulations were conducted to investigate channel hydration and if the differing configurations of substrate binding site residues in the x-ray structures affect substrate binding. Our simulations show that the Q104 side chain does not affect recruitment and binding of NH_4^+ to the substrate binding site, but that it may affect recruitment and binding of other substrates, e.g., MeAmm. We identified novel NH_4^+-protein interactions involved in NH_4^+ recruitment, whereof interactions with D160:O_δ and A162:C=O are of special interest. Our simulations indicate that D160 not only has a structural function, but is also directly involved in NH_4^+ binding at the substrate binding site. A162:C=O was observed to reorient in order to coordinate NH_4^+ when positioned either intraluminally or extraluminally. This suggests that the function of A162:C=O is to guide NH_4^+ across the F107/F215 stack. In addition, we suggest that hydrogen exchange at the A162-G163 peptide bond via the imidic acid mechanism, plausible catalyzed by the D160 carboxylate group and intraluminal NH_4^+, provides a mechanism for transferring a proton from intraluminal NH_4^+ to a periplasmic water molecule. Thus D160 is a prerequisite for conduction.

We have identified two intraluminal binding sites Am1a and Am1b, both providing tetrahedral coordination of NH_4^+, which again stabilizes transient open states of the F107/F215 stack. This indicates that NH_4^+ may translocate from the substrate binding site to intraluminal positions before its deprotonation. Based on our observations and conducted pK_a calculations, we conclude that the protonation state of luminal Histidines prior to any proton

relay is H168:N_δH-H318:N_ϵH. In addition, we propose that the intraluminal binding sites are selective of NH_4^+ over other cations such as K^+ and Na^+, which require different coordination numbers, and thus represent a cation filter. We do not observe intraluminal NH_4^+ translocation across the middle of the pore lumen, which supports the view of NH_3 being the conducted species.

Water is able to enter the channel from cytoplasm via single file formation to H318:N_ϵ(H). Occasionally, a single file can be supported all the way up through the channel lumen. We have implied that our observation of luminal water and single file water may provide a means of proton transfer between H168, H318 and Amm, as well as between lumen and cytoplasm. We do not observe water crossing the periplasmic hydrophobic constriction region. Therefore, we suggest that the F107/F215 stack and the otherwise hydrophobic character of the constriction region constitute a filter against water transport.

Chapter 5

Steered MD simulations of AmtB

Based on: Preliminary results.

5.1 Introduction

It is well established that the *E.coli* ammonia channel AmtB recruits NH_4^+ at the periplasmic substrate binding site (Am1 site) but transports NH_3 across its hydrophobic interior [17, 18, 22, 29, 74, 76, 77, 80]. Therefore, a key question concerning the conduction mechanism of this channel is how and where NH_4 is deprotonated. Previous molecular dynamics (MD) studies have addressed this issue through umbrella sampling simulations seeking to reconstruct free energy profiles as function of the channel axis, i.e., the potentials of mean force (PMFs), for NH_3 and NH_4^+ translocation across AmtB [76, 80]. Although the proposed deprotonation mechanisms differ, both studies essentially agree on the proton being transferred directly to a periplasmic water molecule from NH_4^+ bound at or close to the Am1 site. This contrasts with our previously proposed deprotonation mechanism, where the proton is proposed transferred to A162:O via an imidic acid reaction with NH_4^+ located at a position below the Am1 site (either at the Am1a or Am1b site) where there is no bulk water present (see Chapter 4) [75]. Similar to the umbrella sampling studies we will seek to reconstruct PMFs by means of steered MD (SMD [116, 169]) simulations, but for the part of the translocation pathway delimited by the Am1 and Am2 sites, since deprotonation most probably occur somewhere in-between these sites.

5.2 Modeling, SMD Simulations and PMF Reconstruction

5.2.1 Modeling

Classical molecular dynamics simulations of the *E.coli* AmtB trimer [pdb 1U7G [17]] in a fully hydrated lipid bilayer have been reported elsewhere (see Chapter 4) [75]. As starting point for the present SMD simulations an equilibrated system configuration was extracted from the last part of the 10.0 ns long sim-$\text{III}^{\text{wt}}_{\text{noAmm}}$ simulation. This simulation was started from an initial structure where the two intraluminal Histidines were in the neutral protonation state H168:N_δH–H318:N_ϵH (III), the channel was devoid of any substrate (noAMM $\sim$ no NH_3 nor NH_4^+), and F68 and Q104 were positioned in their wild type conformations (wt). The motivation for conducting SMD simulations was to reconstruct the PMF for translocation of NH_3 and NH_4^+ (Amm) across the F107/F215 stack in AmtB. Consequently, a NH_4^+ ion or a NH_3 molecule was positioned at the Am1 position (see Fig. 5.3B, Z=0 Å) in each monomer of the equilibrated starting structure. In the case of NH_4^+, an appropriate number of bulk water molecules were replaced with chloride ions to ensure overall neutrality. The coordinates of the Am1 positions were determined by superpositioning the Amm bound 1U7G x-ray structure onto each monomer using an rmsd alignment of helical C_α-atoms. Likewise, a water molecule was introduced into the Am2 position (see Fig. 5.3B, Z=−10 Å, exemplified by NH_4^+) of each monomer, since both experimental and computational studies conclude that the channel may be partially hydrated when devoid of substrate [18, 75, 76, 78, 79]. The translocation pathway across AmtB is practically parallel to the z-axis, which was chosen as the reaction coordinate Z with Z=0 Å for the Am1 site (i.e., R=Z in section 3.5).

5.2.2 Simulation Details

Molecular dynamics simulations at constant temperature T=310 K and pressure P=1 atm were conducted using NAMD [105]. Simulation parameters identical to those of the previous study were applied (see Chapter 4) [75]. The systems were first minimized with all protein atoms and Amm nitrogens fixed and subsequently with the protein released. The systems were then equilibrated for 600 ps with the Amm nitrogens harmonically constrained in the z-direction using a force constant of 150 pN/Å. Six equilibrated configurations equally spaced by 100 ps provided the starting structures for the SMD simulations.

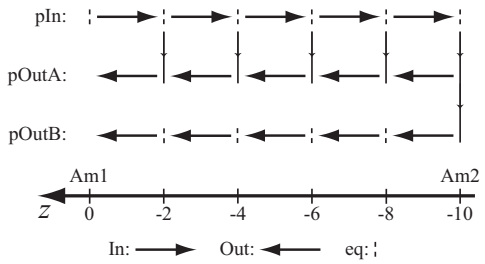

Figure 5.1: The SMD scheme. Schematic drawing of the schemes adopted for the present SMD simulations. Arrows indicate the pulling across segments and the direction; 'In' to or 'Out' of the channel lumen. The reaction coordinate Z spans $\sim$10 Å from the extraluminal Am1 site to the intraluminal Am2 site and is divided into 5 segments of length 2 Å. Equilibration points are indicated by vertical dashed lines. A vertical line with an arrow indicates that the starting point for pulling Out was taken as an equilibrium point of the pulling In simulation. Thus two different schemes pOutA and pOutB were used for the Out direction. In pOutA intermediate equilibration points from pIn are used. In pOutB only the last equilibration point from pIn is used, and thereby it overall resembles the pIn scheme in reverse.

5.2.3 The SMD Scheme

A scheme for conducting SMD simulations was implemented via the Tcl-scripting interface to NAMD. In this scheme harmonic constraints with force constants k and constraint positions $\lambda(t)$ were independently applied to the z-components ($z(\mathbf{q}_t)$) of the center-of-masses of Amm as schematized in Fig. 5.1. The constraint positions were subjected to sequences of *i*) moving with constant velocity v along the reaction coordinate, to pull ("steer/guide") Amm in that direction, and of *ii*) being held fixed, to let the system equilibrate at intermediate λ-values. Such a scheme allows the PMF to be reconstructed independently for each pulling sequence, i.e., segments of the reaction coordinate, thereby minimizing the statistical inaccuracy of the external work [116, 141]. As shown in Fig. 5.1 such a scheme can be realized in two different ways when pulling in both directions (In and Out) along the reaction coordinate. Of course it can be realized as consecutive pulls (pIn and pOutB), where the final equilibrium point serves as the starting point for pulling across the following segment, but it can also be realized as a "parallel" pull (pOutA), where the final equilibrium point of a segment when pulling in one direction is used as the starting point for pulling in the opposite direction across that segment. The latter way allows the pulling simulations in the opposite direction to be run in parallel, thus speeding up the overall simulation process, and at the same time it may also improve the congruence between the oppositely directed pulling experiments.

The part of the reaction coordinate for which the PMF should be reconstructed spans 10 Å from the Am1 position (Z=0 Å) to the Am2 position ($Z\approx-10$ Å, Fig. 5.3B). The lowest possible velocity should be applied in order to be as close to equilibrium as possible during simulation, thereby

minimizing the number of trajectories needed for the PMF reconstruction (see 5.2.5). The velocity determines the length of the trajectory and thereby the computation time. To be able to obtain a reasonable number of trajectories within a reasonable time, we chose to use the velocity v=0.01 Å/ps, which is comparative to what have successfully been applied for other systems [115, 116, 128].

The force constant to be applied is related to the resolution with which the PMF should be reconstructed and to the barrier heights in the PMF. We seek to reconstruct the PMF with a resolution of 1 Å and from experiments the highest barrier has been estimated to 1.6 kcal/mol [34]. According to Eqs. 3.78–3.79 the lower limit of k is given by the requirements k>42 pN/Å and k>220 pN/Å. For the PMF reconstruction process it is required that there is a reasonable agreement between $z(\mathbf{q}_t)$ and λ_t, which in part also is determined by the velocity. Based on several test simulations with different force constants, we found that k=750 pN/Å had to be applied to ensure a reasonable agreement between $z(\mathbf{q}_t)$ and λ_t for v=0.01 Å/ps. This choice of k corresponds to thermal fluctuations along the reaction coordinate of order $\sigma_Z=(\beta k)^{-1/2}\approx$0.2 Å and force fluctuations of order $\sigma_f=(k/\beta)^{1/2}\approx$179 pN with a period $\tau=2\pi/\sqrt{k/m}\approx$0.4 ps for $m=m(\mathrm{NH_4^+})$ (see section 3.5.3 and Eq. 3.83).

The reaction coordinate was divided into 5 segments of length 2 Å (Fig. 5.1). Consequently, to traverse one segment the constraint positions move with constant velocity for 200 ps. It was decided to equilibrate the system for 100 ps between successive pulling sequences. Thus, each SMD simulation was run for 1.5 ns in order to sample the 10 Å of the reaction coordinate. During the SMD simulations Amm were decoupled the temperature bath to avoid artifical forces from the heat bath, and C_α-atoms were harmonically constrained in the z-direction using a force constant of 150 pN/Å. The latter to avoid any translocation of the protein structure as a side-effect of pulling. C_α-atoms of residues 158 to 165 positioned in the irregular helical structure around D160 were not constrained to allow structural rearrangement of back bone atoms, in particular, to allow a possibly critical reorientation of the A162-G163 peptide bond as observed in Chapter 4 [75].

5.2.4 SMD Simulations

Six simulations following scheme pIn (Fig. 5.1), where Amm was pulled into the channel lumen, were conducted to obtain 18 trajectories (Table 5.1, pIn). In all trajectories NH_4^+ entered the channel lumen, while in 9 trajectories NH_3 failed to cross the F107/F215-stack and ended up outside the channel lumen. SMD simulations were also conducted with reversed pulling direction by initially applying scheme pOutA (Fig. 5.1). However, for NH_4^+ a problem occurred using this scheme; displacements in the x, y-plane due to interac-

tions with G163:O caused the final structures in one segment to deviate a lot from the starting structures for the following segment. This lead to discontinuities in the resulting OutA PMF of NH_4^+ (Fig. 5.2, $Z \in [-4; -6]$ and $Z \in [-6; -8]$). Therefore, it was decided to conduct reversed pulling simulations of NH_4^+ following scheme pOutB (Fig. 5.1). This approach also led to problems, e.g., in 5 trajectories NH_4^+ was displaced in the x, y-plane just beneath the F107/F215 stack due to interactions with G163:O and further pulling induced unrealistic structural changes in the nearby protein structure, which precluded the use of these 5 trajectories in subsequent PMF reconstruction.

For NH_3 we only applied scheme pOutA. Although this scheme also lead to discontinuities in the resulting PMFs of NH_3, in this case due to interactions with G163:HN, it was not as severe as for NH_4^+ and therefore no simulations following scheme pOutB were conducted. However, only 3 simulations of NH_3 following scheme pOutA were conducted since, in 3 out of 6 pIn simulations, which should provide the basis for the simulations, NH_3 entered the pore lumen properly only in one of the three monomers and, accordingly, these simulations were discarded.

Amm	Scheme	#Sim.	$\#\text{Traj.}^{\text{Sim}}$	PMF	$\#\text{Traj.}^{\text{PMF}}$	Comments
$\mathbf{NH_4^+}$	pIn	6	18	In	18	
				In-R1	5	
				In-R2	12	
	pOutA	6	18	OutA	18	
	pOutB	6	18	OutB	13	
				OutB-R1	7	Resembles In-R1
				OutB-R2	6	Resembles In-R2
$\mathbf{NH_3}$	pIn	6	18	In	9	
	pOutA	3	9	OutA	6	

Table 5.1: SMD simulations. The SMD simulations of Amm (NH_3 and NH_4^+). The SMD scheme refers to Fig. 5.1. The number of simulations conducted, the corresponding number of trajectories (=3×#Sim.), and the number of trajectories used for PMF reconstruction are listed. "R1" and "R2" denote two different routes of NH_4^+ translocation (see text).

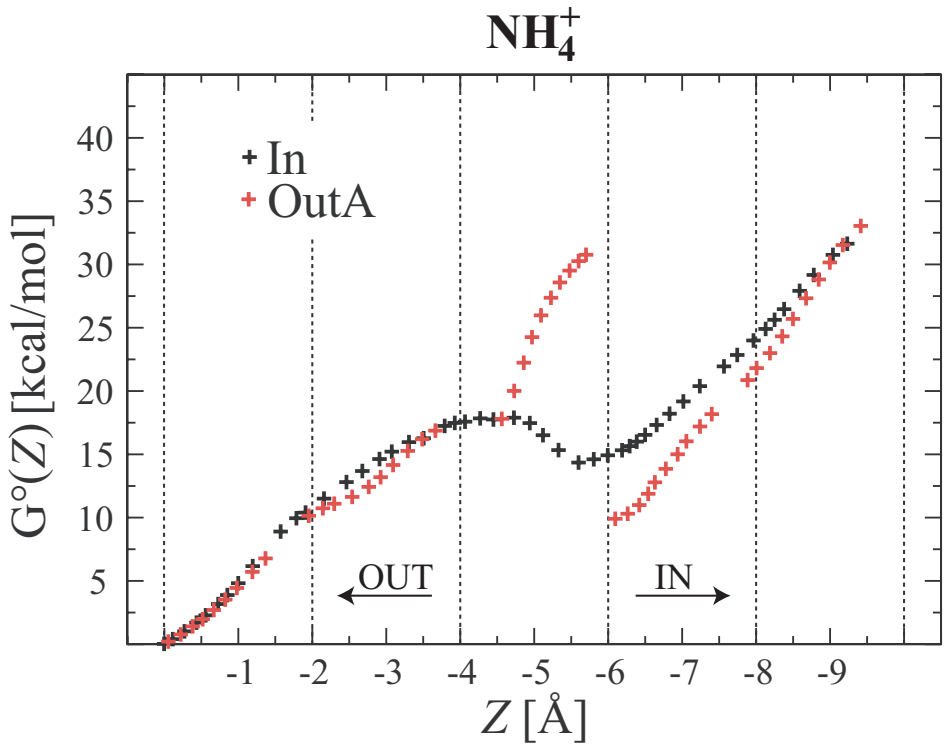

Figure 5.2: PMFs for NH_4^+. PMFs reconstructed from the pIn and pOutA simulations of NH_4^+. The reconstruction procedure is described in section 5.2.5. Discontinuities in the simulations following scheme pOutA lead to discontinuities in the corresponding OutA PMF in segment $Z \in [-4; -6]$.

Average	Procedure of calculation
$\overline{t}(j)$	$\frac{1}{\Delta t}\int_j \mathrm{d}t\ t$
$\overline{z}_i(j)$	$\frac{1}{\Delta t}\int_j \mathrm{d}t\ z_i(\mathbf{q}_t)$
$\langle\overline{z}(j)\rangle_F$	$\frac{1}{N}\sum_{i=1}^{N}\ \overline{z}_i(j)$
$\overline{V}(j)$	$\frac{k}{2}\left[\langle\overline{z}(j)\rangle_F - \lambda_0 - v\overline{t}(j)\right]^2$
$\overline{\mathcal{W}}_i(j)$	$\frac{1}{\Delta t}\int_j \mathrm{d}t\ \mathcal{W}_i(t)$
$\overline{\mathscr{W}}_i(j)$	$\overline{\mathcal{W}}_i(j) - \overline{V}(j)$
$\langle\overline{\mathscr{W}}(j)^m\rangle_F$	$\frac{1}{N}\sum_{i=1}^{N}\ \overline{\mathscr{W}}_i(j)^m$

Table 5.2: Time and ensemble averages. j denotes the time window ($j \in [1,2,\ldots,K]$). i denotes the trajectory ($i \in [1,2,\ldots,N]$). The time average of a quantity x is denoted by $\overline{x}$, while the ensemble average is denoted by $\langle x\rangle_F$. The integral $\int_j$ refers to the integral in Eq. 3.80 and hence denotes a time average over the jth time window. $z_i(\mathbf{q}_t)$ is the z-component of the c.o.m. of the substrate with coordinates $\mathbf{q}_t$ at time t in the ith trajectory. The cumulated external work $\mathcal{W}_i(t)$ to time t of the ith trajectory was calculated using Eq. 3.58.

5.2.5 PMF Reconstruction

Details concerning the theoretical background for PMF reconstruction were provided in Chapter 3. Here we apply the approach followed by Jensen et al. [116], by reconstructing the PMF independently in each segment using Eq. 3.85 expanded into cumulants as in Eq. 3.37 and by assuming that $\langle\overline{r}(j)\rangle = \langle\overline{r}(j)\rangle_F$ (here r=z), which is the time averaged analog of Eq. 3.75. At NPT conditions Eq. 3.85 then becomes [147]

$$G^{\circ}(\langle\overline{z}(j)\rangle_F)_{\overline{\lambda}(j)} \approx \left\langle\overline{\mathscr{W}}(j)\right\rangle - \frac{\beta}{2}\left[\left\langle\overline{\mathscr{W}}(j)^2\right\rangle - \left\langle\overline{\mathscr{W}}(j)\right\rangle^2\right], \qquad (5.1)$$

The time and ensemble averaged quantities entering this equation were calculated as summarized in Table 5.2.

The time averaging were done in K windows of length 20 ps each encompassing $\frac{20\,\mathrm{ps}}{\tau}\approx$50 periods of the thermal fluctuations. A window of length 40 ps was also applied to test the sensitivity of the PMF with respect to window size. No significant sensitivity was found. The sectional PMFs were connected by *1*) fitting a second order polynomial to the first three data points in a segment and *2*) adjusting this segment such that an extrapolation of the polynomial coincided with the last data point of the previous segment. This procedure was chosen, since the first data points are the best estimated. To obtain the smoothened PMFs in Fig. 5.3A a Fourier series composed of M (=10) sine functions and M cosine functions, $\tilde{G}(Z) = \sum_{m=1}^{M} a_m \sin[m\pi(Z-z_1)/(z_K-z_1)] + b_m\cos[m\pi(Z-z_1)/(z_K-z_1)]$, was fitted to all data points between the first one (z_1=$\langle\overline{z}(1)\rangle_F\approx$0 Å) and the last one ($z_K$=$\langle\overline{z}(K)\rangle_F\approx-$10 Å). All In PMFs were then adjusted to

satisfy the condition $G^{\circ}(0)$=0 (Figs. 5.2 and 5.3A), except the In PMF of NH_3 (Fig. 5.3A, right panel), which was adjusted according to an estimated relative binding free energy (see Results and Discussion). The OutA and OutB PMFs were adjusted manually such that the PMFs matched the final part (Z≈−10 Å) of the corresponding In PMFs (Figs. 5.2 and 5.3A, left panel). Due to the discontinuities in the middle segment ($Z \in [-4;-6]$) of the OutA PMF, this was also matched at Z≈−4 Å (Fig. 5.2).

5.2.6 Accuracy of the PMFs

Hysteresis in the PMFs reconstructed from the pIn and pOut simulations, as well as large work fluctuations $\sigma_{\mathcal{W}_{total}}$, can be indicative of a low accuracy of the calculated PMFs.

By comparing PMFs reconstructed from inward and outward pulling of NH_4^+ (Fig. 5.2, In vs. OutA PMFs) we find that hysteresis is low in two segments ($Z \in [0;-4]$), while the other segments show some degree of hysteresis, especially segment $Z \in [-4;-6]$. However, as mentioned above the latter can be attributed to the noncoinciding final and initial structures between segments. When comparing the In PMF to the PMFs reconstructed for outward pulling following scheme pOutB [Fig. 5.3A, In vs. OutB(-R1/R2) PMFs in left panel] a large degree of hysteresis is seen in the first three segments ($Z \in [0;-6]$), except for the In-R1 and OutB-R1 PMFs which only show hysteresis in the first two segments ($Z \in [0;-4]$). Seemingly, the hysteresis in these PMFs can at least partly be attributed to the final structures of the pOutB simulations, which did not resemble the initial structures of the pIn simulations, e.g., the F107/F215-stack did not obtain its closed conformation as in Fig. 5.3B (Z=0 Å). Since the OutB PMFs were made to match the corresponding In PMFs at Z≈−10 Å, this camouflages that some of the hysteresis may be assigned to this part of the PMFs, i.e, the OutB PMFs should be positioned at a lower position, thus resulting in lower barriers of at least the In/OutB and the In-R1/OutB-R1 PMFs. Comparing the In and OutA PMFs for NH_3 revealed almost no hysteresis.

For each segment in the PMFs we calculated the standard deviation $\sigma_{\mathcal{W}_{total}} = \sqrt{\langle \mathcal{W}_{total}^2 \rangle - \langle \mathcal{W}_{total} \rangle^2}$ (Table 5.3) of the total external work $\mathcal{W}_{total}$, which is related to the total dissipative work as $\mathcal{W}_{diss,\,total} = (\beta/2)\sigma^2_{\mathcal{W}_{total}}$. Hence $\sigma_{\mathcal{W}_{total}}$ is as a measure of the the nonequilibrium character of the simulations and consequently grows with increasing pulling speed. For direct application of the Jarzynski equality $\sigma_{\mathcal{W}_{total}}$ should be less than a few k_BT [140], and when resorting to the second order cumulant expansion, it should also be relatively small to obtain converged estimates of $\langle \mathcal{W} \rangle$ and $\langle \mathcal{W}^2 \rangle$ from few trajectories [115]. For PMFs reconstructed using the stiff spring approximation good accuracy has been obtained for $\sigma_{\mathcal{W}_{total}}$≈3 kcal/mol, while poorer accuracy occur for $\sigma_{\mathcal{W}_{total}}$≈7 kcal/mol [115]. Though the re-

construction method applied here is different, both methods resort to the second order cumulant expansion. Therefore, we assume that $\sigma_{\mathcal{W}_{total}}$ should not be much larger than[1] 3 kcal/mol in order to reconstruct accurate PMFs. Table 5.3 lists $\sigma_{\mathcal{W}_{total}}$ calculated for the sets of trajectories corresponding to the reconstructed PMFs. According to these values and in correspondance with the observed degree of hysteresis above, the most accurate PMFs are obtained for NH_3. Also the In, In-R1 and In-R2 PMFs of NH_4^+ can be expected to be reasonably accurate except perhaps for the middle segment ($Z \in [-4; -6]$), while all the 'Out' PMFs of NH_4^+ have relatively large $\sigma_{\mathcal{W}_{total}}$ values.

The above considerations indicate that the PMFs most accurately representing the energetics of NH_4^+ translocation are the In PMFs, which also happens to be reconstructed from simulations of the physiological direction of Amm transport. For NH_3 both the In and Out PMFs are representative of the energetics. Therefore, the focus of the following section will be on these PMFs.

Amm	PMF	$\sigma_{\mathcal{W}_{total}}/[k_B\mathrm{T}]$				
		$Z \in$ [0; −2]	[−2; −4]	[−4; −6]	[−6; −8]	[−8; −10]
NH_4^+	In	2.6	4.3	**9.4**	4.4	3.9
	In-R1	3.1	4.4	**6.2**	**5.1**	2.0
	In-R2	2.5	4.4	**6.8**	4.0	4.5
	OutA	2.7	**5.5**	**11.4**	**6.3**	**6.9**
	OutB	**5.4**	**6.6**	**6.5**	**6.4**	**7.4**
	OutB-R1	**6.7**	**7.2**	**7.0**	5.0	**8.2**
	OutB-R2	3.2	**5.9**	3.6	**7.3**	**6.3**
NH_3	In	3.2	1.9	3.7	**5.3**	2.7
	OutA	3.5	1.9	4.4	3.4	1.7

Table 5.3: Work fluctuations. The standard deviation $\sigma_{\mathcal{W}_{total}}$ of the total work performed within each pulling segment is listed (in k_BT) for the sets of trajectories corresponding to the reconstructed PMFs. Values larger than $5.0k_B$T are in bold.

[1] We are aware that this value also depends on the nature of the simulated problem.

5.3 Results and Discussion

5.3.1 NH_4^+ Follows Two Routes Into Pore Lumen

In the pIn simulations NH_4^+ follows two distinctive routes across the F107/F215-stack (Fig. 5.3B, In-R1 and In-R2). At first the two routes coincide and no differences in the corresponding PMFs can be observed (Fig. 5.3A–B, Z=[0, −2]), but the level of NH_4^+ hydration differ. Around Z=−2 Å NH_4^+ generally coordinates 1–2 waters in In-R1 and 2–3 waters in In-R2. Further down towards the lumen the two routes start to spatially diverge (Fig. 5.3B, Z=[−4, −6]). In In-R1 NH_4^+ coordinates directly to F215:O and N216:O_δ while in In-R2 NH_4^+ is positioned more back left and through water molecules coordinates to F107:O and occasionally to F215:O. For In-R1 there is a local minimum in $G°(Z)$ in this part of the reaction coordinate, while no minimum is observed in In-R2. This is what we would expect, since In-R1 is the only route, which passes through our previously identified intraluminal NH_4^+ binding sites (the Am1a and Am1b sites in Chapter 4). Below the F107/F215-stack the two routes spatially coincides again (Fig. 5.3B, Z=[−8, −10]), however, the In-R2 PMF is higher in energy than In-R1 possibly due to different levels of hydration inside the lumen; 0–1 waters are co-transported in In-R1, while 1–2 in In-R2. The In-R1 PMF is lowest in energy and has the lowest barrier and should therefore be the preferred route for NH_4^+ to follow. The low sampling of route In-R1 in comparison to In-R2 (Table. 5.1) can be due to at least two effects 1) a too high pulling velocity and/or 2) a non-optimal pulling direction. To test if the samling of In-R1 could be improved by reducing the velocity, one simulation was conducted with half velocity, but was inconclusive in this respect. Although a reduced velocity in this case might not improve the sampling efficiency much, it is likely to decrease the dissipative work expended in the middle segment and thereby increase the accuracy of the reconstructed PMF. Thus a reduced velocity combined with adjustments of the pulling direction to more directly guide NH_4^+ along In-R1, are subjects for future studies.

5.3.2 A162 Guided Ammonium Transfer

As observed in Fig. 5.3B, A162:C=O coordinates to NH_4^+ all the way from the Am1 site (Z=0 Å) to an intraluminal position (Z≈−6 Å). The ability of A162:C=O to coordinate NH_4^+ at the Am1 position, as well as to reorient from intra- to extraluminal positions, supports its important role in NH_4^+ stabilization [74–76], and in recruiting and guiding NH_4^+ across the F107/F215 stack (see Chapter 4) [75].

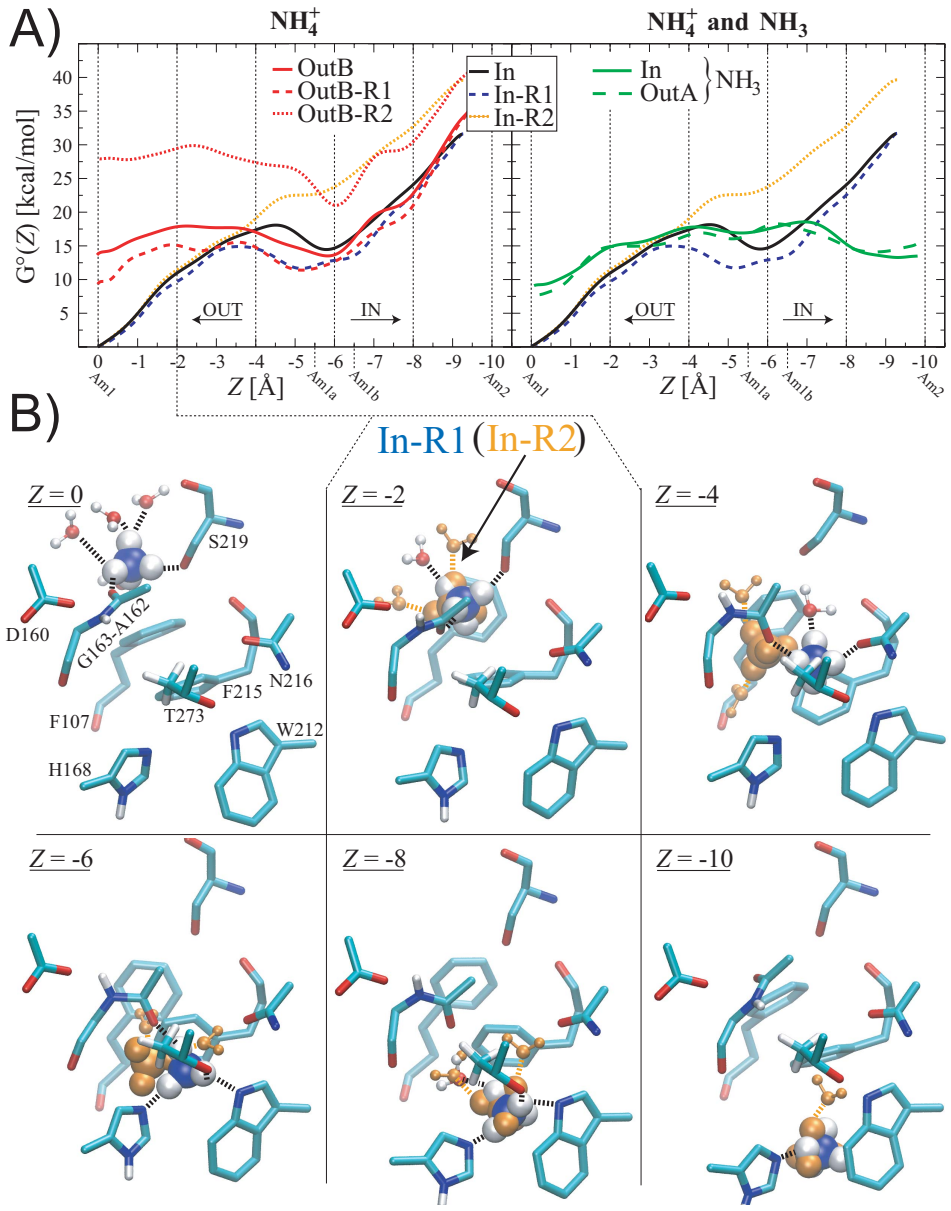

Figure 5.3: PMFs for NH_4^+ and NH_3. (A) PMFs reconstructed from the pIn simulations of NH_4^+ are included on both the left and right panels. In-R1 and In-R2 are PMFs corresponding to the two routes followed across the F107/F215-stack as visualized in B. PMFs reconstructed from the pOutB simulations of NH_4^+ are shown on the left panel, where OutB-R1 and OutB-R2 are PMFs essentially corresponding to the reverse of routes In-R1 and In-R2. PMFs reconstructed from the pIn and pOutA simulations of NH_3 are shown on the right panel, (B) Representative snapshots of equilibrated In-R1 structures at the initial (Z≈0 Å), intermediate (Z≈[−2, −4, −6, −8]) and final (Z≈−10 Å) equilibration points. Key residues (*licorice*), NH_4^+ (*vdW*) and coordinating water (*ball-and-stick*) are shown and colored according to atom type. Except at Z≈0 Å, where In-R1 and In-R2 are coinciding, representative snapshots of NH_4^+ (*vdW, orange*) and coordinating water (*ball-and-stick, orange*) in the In-R2 structures are also shown.

5.3.3 Deprotonation of NH_4^+

As NH_4^+ moves from bulk water and into the pore lumen the environment becomes hydrophobic and at some intermediate point it will be energetically favorable to deprotonate and become NH_3. The driving force for that event is the Gibbs free energy change associated with the deprotonation reaction. Below is outlined a procedure, which allow us to localize the region of the channel where NH_4^+ is most likely deprotonated.

The dissociation of NH_4^+ at any point Z along the reaction coordinate obeys the acid-equilibria

$$\mathrm{NH_4^+}\,(Z) \leftrightarrows \mathrm{NH_3}\,(Z) + \mathrm{H^+}\,(Z) \quad , \quad \mathrm{pK}_a(Z) = -\log\left(\frac{\mathrm{a(NH_3};Z)\cdot \mathrm{a(H^+};Z)}{\mathrm{a(NH_4^+};Z)}\right), \tag{5.2}$$

where a(X;Z) is the activity of molecule X at position Z. In bulk water $\mathrm{pK}_a(aq) = 9.25$, which makes NH_4^+ the predominant species at pH=7 ($\alpha = \mathrm{a(NH_4^+};aq)/\mathrm{a(NH_3};aq) \approx 178$). Deprotonation of NH_4^+ is most likely to occur in the region where both species are equally concentrated ($\alpha = 1$, the equivalence point), which will happen when $\mathrm{pK}_a(Z)$=pH(Z) implying

$$\mathrm{pK}_a(Z) - \mathrm{pH}(Z) = -\log\left(\frac{\mathrm{a(NH_3};Z)}{\mathrm{a(NH_4^+};Z)}\right) = 0. \tag{5.3}$$

To determine the deprotonation region we therefore need to know $\mathrm{pK}_a(Z)$ and pH(Z). $\mathrm{pK}_a(Z)$ can be calculated from a so-called pK_a-shift calculation as outlined below. $\mathrm{pK}_a(Z)$ is related to the standard free energy of the reaction as

$$\begin{aligned}\mathrm{pK}_a(Z) &= \frac{\Delta G_a^\circ(Z)}{\mathrm{RTln}\,10} = \frac{G^\circ(\mathrm{NH_3};Z) + G^\circ(\mathrm{H^+};Z) - G^\circ(\mathrm{NH_4^+};Z)}{\mathrm{RTln}\,10} \\ &= \frac{\Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}} G^\circ(Z) + G^\circ(\mathrm{H^+};Z)}{\mathrm{RTln}\,10}\end{aligned} \tag{5.4}$$

where $\Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}} G^\circ(Z)$ is the mutational standard free energy for the mutation $\mathrm{NH_4^+}(Z) \to \mathrm{NH_3}(Z)$. The difference $\mathrm{pK}_a(Z)-\mathrm{pK}_a(aq)$ can then be expressed as

$$\begin{aligned}&\mathrm{pK}_a(Z) - \mathrm{pK}_a(aq) = \\ &\qquad \frac{\Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}} G^\circ(Z) - \Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}} G^\circ(aq)}{\mathrm{RTln}\,10} + \frac{\Delta_{aq\to Z} G^\circ(\mathrm{H^+})}{\mathrm{RTln}\,10}\end{aligned} \tag{5.5}$$

where $\Delta_{aq\to Z} G^\circ(\mathrm{H^+})$ is the standard binding free energy for the process $\mathrm{H^+}(aq) \to \mathrm{H^+}(Z)$. The numerator in the first term on the right hand side is related to the relative binding free energy for binding NH_4^+ vs. NH_3

[shorthand $\Delta\Delta_{aq\to Z}G^\circ(\mathrm{NH_4^+},\mathrm{NH_3})$] as

$$\begin{aligned}\Delta\Delta_{aq\to Z}G^\circ(\mathrm{NH_4^+},\mathrm{NH_3}) &= \Delta_{aq\to Z}G^\circ(\mathrm{NH_4^+}) - \Delta_{aq\to Z}G^\circ(\mathrm{NH_3}) \\ &= -\left[\Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}}G^\circ(Z) - \Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}}G^\circ(aq)\right].\end{aligned} \tag{5.6}$$

The mutational standard free energy may in principle be calculated from alchemical simulations at different points along Z (here the free energy PMFs of $\mathrm{NH_4^+}$ and $\mathrm{NH_3}$ provide the same information when positioned mutually correct, see below). Contrary, does a calculation of $\Delta_{aq\to Z}G^\circ(\mathrm{H^+})$ in Eq. 5.5 pose a problem, since it is not known if $\mathrm{H^+}$ at position Z is bound to the protein and if so where? We therefore have to make an assumption about the magnitude of $\Delta_{aq\to Z}G^\circ(\mathrm{H^+})$. In standard pK_a-shift calculations one assumes that the proton is delivered from/to bulk water according to the equilibria $\mathrm{AH}\,(Z) \leftrightarrows \mathrm{A^-}\,(Z) + \mathrm{H^+}\,(aq)$, where A is the conjugated base [170], which implies that $\Delta_{aq\to Z}G^\circ(\mathrm{H^+}) = 0$. Here an assumption having the same effect would be to assume that the standard chemical potential of $\mathrm{H^+}$ at position Z is almost the same as that in bulk water, i.e., $G^\circ(\mathrm{H^+},Z) \approx G^\circ(\mathrm{H^+},aq)$, such that the second term in Eq. 5.5 is small compared to the first term and hence can be neglected. For a system in equilibrium holds

$$\Delta_{aq\to Z}G(\mathrm{H^+}) = \Delta_{aq\to Z}G^\circ(\mathrm{H^+}) + \mathrm{RT}\ln\left[\frac{\mathrm{a}(\mathrm{H^+};Z)}{\mathrm{a}(\mathrm{H^+};aq)}\right] = 0 \tag{5.7}$$

and consequently the assumption involve that $\mathrm{a}(\mathrm{H^+};Z) = \mathrm{a}(\mathrm{H^+};aq)$. As a consequence we will have that $\mathrm{pH}(Z)$=7, which clearly is questionable if position Z is inside a protein; if locally $\mathrm{a}(\mathrm{H^+}) = 0$ then pH should be infinitely large! Hence, the assumption is probably valid at the protein-solvent interface, where there is direct contact to bulk water, but otherwise it is encumbered with uncertainty. Additionally, the validity of such an assumption can be questioned for Amm, especially if a change in $\Delta_{aq\to Z}G^\circ(\mathrm{NH_4^+})$ can mainly be attributed to electrostatic effects, since this should invoke a change in $\Delta_{aq\to Z}G^\circ(\mathrm{H^+})$ of the same magnitude, which might not render it negligible. However, following the above procedure to locate the deprotonation region of $\mathrm{NH_4^+}$ allows a direct comparison to the results of a previous study on AmtB [80]. Thus, we assume that the standard chemical potential of $\mathrm{H^+}$ at position Z is almost the same as that in bulk water, i.e., $G^\circ(\mathrm{H^+},Z) \approx G^\circ(\mathrm{H^+},aq)$, such that the second term in Eq. 5.5 is small compared to the first term and hence can be neglected. The PMF reconstruction procedure applied here actually yields PMFs $\delta G^\circ(\mathrm{X};Z) = G^\circ(\mathrm{X};Z) - G^\circ(\mathrm{X};\mathrm{Am1})$ reflecting the free energy of X relative to that at the initial value of the reaction coordinate (note that in the figures δG° is written as G°). The difference between the reconstructed PMFs $\delta G^\circ(\mathrm{NH_3};Z)$ and $\delta G^\circ(\mathrm{NH_4^+};Z)$ then reads

$$\Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}}\delta G^\circ(Z) = \Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}}G^\circ(Z) - \Delta_{\mathrm{NH_4^+}\to\mathrm{NH_3}}G^\circ(\mathrm{Am1}) \tag{5.8}$$

Substituting this into Eq. 5.5 (where the second term is neglected) yields upon rearrangement

$$\begin{aligned}\Delta_{\mathrm{NH_4^+}\rightarrow\mathrm{NH_3}}\delta G^\circ(Z) = [\mathrm{pK}_a(Z) - \mathrm{pK}_a(aq)] \cdot \mathrm{RTln}\,10 \\ + \Delta\Delta_{aq\rightarrow\mathrm{Am1}}G^\circ(\mathrm{NH_4^+}, \mathrm{NH_3}),\end{aligned} \tag{5.9}$$

where the definition of the relative binding free energy in Eq. 5.6 has been utilized. From a previous computational study of AmtB [74] we have that

$$\Delta\Delta_{aq\rightarrow\mathrm{Am1}}G^\circ(\mathrm{NH_4^+}, \mathrm{NH_3}) = -6 \text{ kcal/mol} \tag{5.10}$$

The deprotonation region [$\mathrm{pK}_a(Z)$=7] is then centered where the difference between the two PMFs is

$$\begin{aligned}\Delta_{\mathrm{NH_4^+}\rightarrow\mathrm{NH_3}}\delta G^\circ(Z) = [7 - 9.25] \cdot \mathrm{RTln}\,10 - 6 \text{ kcal/mol} \\ \approx -9 \text{ kcal/mol}, \quad \mathrm{T} = 310 \text{ K}.\end{aligned} \tag{5.11}$$

However, since it will be easier to read off the deprotonation point as where the two PMFs intersect we instead position the NH_3 PMF 9 kcal/mol above the NH_4^+ PMF at the Am1 site.

In Fig. 5.3B the PMF reconstructed for the NH_3 pIn simulations is positioned above the NH_4^+ PMFs at the Am1 site in correspondance with the procedure outlined above. At this site NH_4^+ is in direct contact with water in the periplasmic vestibule and the assumption of $\Delta G^\circ_{aq\rightarrow\mathrm{Am1}}(\mathrm{H^+})$ being small compared to $\Delta\Delta G^\circ_{aq\rightarrow\mathrm{Am1}}(\mathrm{NH_4^+}, \mathrm{NH_3})$ and of pH(Am1)=7 are presumably reasonable, whereby $\mathrm{pK}_a(\mathrm{Am1}) = 13.6$. This precludes the existence of NH_3 at this position ($\alpha \approx 4 \times 10^6$). The two PMFs intersect at $Z{\approx}{-}7$ Å and NH_4^+ should therefore deprotonate somewhere around this position. Although, the underlying assumptions might be less valid at this position the result suggests that NH_4^+ is completely disconnected from periplasmic water when it deprotonates, which is in contrast with a previous MD study [80], where NH_4^+ was suggested to deprotonate directly to periplasmic water at a position corresponding to $Z{=}{-}2$ Å. In another MD study it was also suggested that NH_4^+ deprotonates directly to periplasmic water (and eventually to D160) [76]. However, this suggestion was based on an arbitrary superpositioning of the NH_4^+ and NH_3 PMFs at the Am1 site. Instead, if separating their PMFs by 9 kcal/mol at the Am1 site the deprotonation region moves several Ångström below, which seems to agree with our results. Previously, we have suggested a deprotonation mechanism for NH_4^+ when it is disconnected from periplasmic water (see Chapter 4) [75]. This mechanism was rejected by Bostick and Brooks III [80], but according to our results this seems like a premature conclusion.

5.3.4 Barrier Against Transport

The process of moving NH_4^+ from the Am1 site and into the deprotonation region is associated with a barrier of height ∼15 kcal/mol (Fig. 5.3A,

right panel, In-R1), which is way too high in comparison to the experimentally determined activation energies $E_A = 1.6$ kcal/mol for MeAmm transport across AmtB [34] and $E_A = 11.8 \pm 2.8$ kcal/mol for NH_3 permeability through cell membranes [171]. The barrier heights reported in the two other MD studies 6.9 kcal/mol [76] and 8.9 kcal/mol [80] are therefore also too high. The reason why our barrier is higher than these two can be due to an insufficient sampling along In-R1 (Table 5.1), which combined with the large work fluctuations in the barrier region (Table 5.3, $Z \in [-4; -6]$), are introducing uncertainty of the PMF in this region. However, this does not affect the above conclusion that NH_4^+ is disconnected from periplasmic water when it deprotonates, since the NH_3 In and OutA PMFs apparently are well estimated (show almost no hysteresis and the standard deviation of the external work is low) and following an improved NH_4^+ PMF with lower barrier would probably intersect at $Z \lesssim -7$ Å implying that the deprotonation region moves even further away from the periplasmic vestibule.

In all three MD studies, the position of the barrier corresponds to the crossing of the F107/F215-stack. Previously, AmtB has been suggested to be a slowly conducting channel with a transport rate of only 10–10000 molecules/s ($t_{transport} > 1 \times 10^5$ ns/molecule) [18, 34], with a thermally activated opening of the F107/F215-stack being the limiting factor [18]. Hence, the large barrier heights resulting from the computational studies might be due to a forced opening of the F107/F215-stack. Another explanation could be an incorrect representation of cation-π interactions [160], which are important to get right, since NH_4^+ gets in close encounter with F107 and F215, when passing the F107/F215-stack. To address this issue future investigations based on *ab initio* methods are required. Other effects which can provide some explanation are *i*) nonequilibrium coordination of Amm due to the short time scale of conducted MD simulations, *ii*) insufficient protein relaxation, *iii*) inadequate sampling of distinct monomer conformations, *iv*) more than one Amm molecule involved in the translocation mechanism. To address these issues more and longer (slower pulling velocity) trajectories simulated by classical MD would be required.

5.4 Conclusion

By means of steered molecular dynamics simulations PMFs were reconstructed for translocation of NH_3 and NH_4^+ from the Am1 site, across the F107/F215-stack and into the intraluminal Am2 site, as well as for the reverse process. In the simulations NH_4^+ followed two different routes across the F107/F215-stack, one with low hydration and 0–1 co-transported water molecules and another with more hydration and more co-transported water molecules. Since the latter route is higher in energy this is believed not to be physiological relevant. Utilizing a previously calculated relative binding free

energy for binding NH_4^+ vs. NH_3 at the Am1 site [74], allowed us to identify the region where NH_4^+ deprotonates and to compare with two previous MD studies [76, 80]. We find that NH_4^+ should deprotonate when almost past or past the F107/F215-stack, where it is disconnected from periplasmic water. This precludes that the proton is transferred directly to periplasmic water molecules as some have proposed [76, 80], but is consistent with our previously proposed proton transfer mechanism (see Chapter 4) [75]. However, we find that the interpretation of the results of Lin et al. is erroneous due to an arbitrary positioning of their PMFs, and when corrected seemingly agree with our results. In all these computational studies, including the present one, the computed barrier against transport are in comparison to experimental values [34, 171] overestimated. This indicates that all three studies might suffer from insufficient sampling and nonequilibrium conditions; in the present study, hysteresis and large work fluctuations indicated that the applied SMD scheme could be refined to obtain more accurate PMFs especially with respect to NH_4^+. Another explanation of the high barriers, could be an inadequate representation of cation-π interactions [160]. The latter suggests that other computational methods, e.g., *ab initio* methods, needs to be applied in future studies to unravel exactly how and where NH_4^+ deprotonates.

Chapter 6

Hybrid QM/MM MD Simulation of AmtB

Adapted from: *Ab Initio* Molecular Dynamics Study of Substrate Recognition and Cation-π Interactions in *E.coli* Ammonia Channel AmtB. Nygaard, T. P., G. H. Peters, and M. Ø. Jensen, and C. Rovira. *To be submitted.*

6.1 Introduction

Ammonia/ammonium (Amm) is an essential growth factor for some bacteria, yeasts, plants, and fungi, while for other organisms it is a highly toxic metabolic waste product, e.g., for mammalian cells [1, 43]. Proteins belonging to the ammonia/ammonium transporter (AMT) family have been identified in a wide range of organisms; (methyl-)ammonia/ammonium permeases (Mep) in yeasts, Rhesus (Rh) blood group proteins in animals, ammonia/ammonium transporter (Amt) proteins in bacteria and plants [8, 9]. Therefore, a detailed understanding of how this family of proteins function may have future pharmaceutical, agricultural , and biotechnological relevance [15, 68, 69, 72].

The first AMT protein for which high resolution crystallographic structures were solved was *E.coli* AmtB [17, 18], followed by *Archaeoglubus fulgidus* Amt-1 [20], and *E.coli* AmtB complexed with its physiological inhibitor GlnK [21, 22]. Three identical AmtB monomers form a trimer [17, 18, 70, 71]. Each monomer has eleven transmembrane-spanning α-helices arranged in a right-handed bundle. Depressions in both the periplasmic and the cytoplasmic surfaces (vestibules) lead into a hydrophobic pore lined by Histidines H168 and H318 (Fig. 6.1A–B,D), which, by sharing a H-atom between their imidazole rings, are mutually fixed [17, 18]. In the crystal structures of Khademi et al. a periplasmic $NH_4^+/MeNH_4^+$ molecule and three intraluminal gaseous NH_3 molecules were resolved and their positions were named Am1, Am2, Am3, and Am4, respectively [17]. A subse-

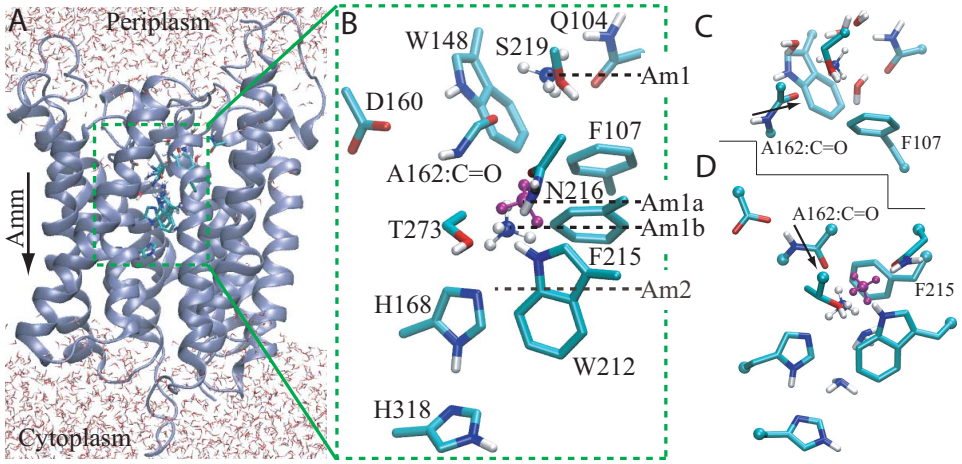

Figure 6.1: The simulated system. (A) The system of the present CPMD simulations. The arrow indicate the direction of physiological Amm transport. (B) Enlargement of important substrate binding site residues and channel lining residues treated at the QM level in the simulations. NH_4^+ is shown at the Am1, Am1a (*purple*), and Am1b positions, while the Am2 position is only indicated. The structure is the starting structure for the Am1 simulations. (C) The QM fragment chosen for the Am1 simulations and (D) for the Am1a and Am1b simulations (see also Table 6.2). The arrows indicate the different orientations of A162:C=O for different NH_4^+ positions. Link atoms are represented by spheres (*cyan*).

quent molecular dynamics (MD) study of AmtB identified two intermediate positions; Am1a and Am1b between Am1 and Am2 (Fig. 6.1B), these may accommodate NH_4^+ (see Chapter 4) [75]. Even though NH_3 recognition at the Am1 site has been proposed [29], the most common view is that this site recognizes NH_4^+ [17, 18, 34]. The suggestion that NH_3 occupies the channel lumen is in accordance with experimental AmtB data, which consistently indicate gradient driven NH_3 uniport, i.e., diffusion or passive transport [17, 29, 34]. That NH_4^+ is the Amm species preferred at the Am1 site and that NH_3 occupies the Am2–Am3 sites is also supported by computational studies [74, 77, 80].

The most important residues located in the vicinity of the periplasmic substrate binding site Am1 are Q104, F107, W148, D160, A162, and S219 (Fig. 6.1B). In the vicinity of the intraluminal positions Am1a, Am1b and Am2 the most important residues are D160, A162, H168, W212, N216, and T273 (Fig. 6.1B). The carbonyl oxygen of A162 is hereafter referred to as A162:C=O. With respect to substrate positions, the function(s) of these residues suggested in crystallographic and computational studies are summarized in Table 6.1.

Am1	
Residue(s)	Function(s)
Q104	NH_4^+ recognition due to steric exclusion effects [75].
F107 and W148	Stabilization and selection of charged substrates by cation-π interactions [17, 73, 76, 78]. A gate, which can open and close spontaneously [75, 78] or in correlation to NH_4^+ contact [75, 76].
D160	Electrostatic stabilization of positively charged substrate [74, 75, 77] and/or stabilization of protein structure [17, 18, 75, 80]. Proton acceptor in deprotonation pathway via two-water wire [76].
A162	Stabilization of charged substrates [18, 74–76], dynamically involved in translocating substrate across the F107/F215 stack [75].
S219	Stabilization of charged substrates [17, 18, 75–77]. Catalyzes the deprotonation of NH_4^+ [77].
Am1a, Am1b and Am2	
Residue(s)	Function(s)
D160	Proton acceptor in deprotonation pathway via the A162-G163 peptide bond [75].
A162, H168, W212, N216 and T273	Tetrahedral coordination of NH_4^+ [75, 76].
W212 and F215	Stabilization of charged substrates by cation-π interactions [17, 76].
H_2O and NH_3	Can provide stabilization of NH_4^+ [75, 79].

Table 6.1: Assigned Functions to Residues. The functions of residues with respect to substrate positioned at the Am1 site or at the Am1b, Am1a and Am2 sites, as suggested in crystallographic and computational studies.

AmtB has been the focus of several recent computational studies [17, 18, 73–80], but still many questions and controversies regarding the detailed mechanism of AmtB prevail. In particular, the relative importance of the residues of Table 6.1 with respect to substrate binding and recognition, translocation and deprotonation is still controversial. One example is cation-π interactions, which have been proposed to be essential for binding NH_4^+ to the Am1 and/or Am2 site [17, 73, 78]. However, the importance of cation-π at the Am1 and/or Am2 site has been questioned [74, 75, 77]. Another example is proton transfer mechanisms. Previously we have proposed how an intraluminally positioned (dehydrated) NH_4^+ may deprotonate via the A162-G163 peptide bond (see Chapter 4) [75], but deprotonation just below the the Am1 position via a two-water wire to the D160 carboxylate group [76], deprotonation to periplasmic water molecules, either directly [80] or catalyzed by S219 [77], have also been proposed. Proton transfer via the two-water wire mechanism is, however, highly unlikely since both NH_4^+ at the Am1 position and D160 are partially solvent exposed. Thus, one would not expect their mutual difference in pK_a values to change significantly from their difference in bulk water, where the pK_a of NH_4^+ (9.25) is larger than that of Aspartate (3.7). This argumentation is supported by

recent pK_a-shift calculations, which yielded ΔpK_a values in the range -0.3 to -5.1 for D160 when the Am1 site is occupied by water, and -9.2 when occupied by NH_4^+ [74]. Therefore, if NH_4^+ deprotonates before it is completely dehydrated, a periplasmic water molecule rather than D160 will be the final proton acceptor.

Since the function of cells crucially rely on differences in ion concentrations between the inside and outside, AMT proteins differentiate between NH_4^+ and, e.g., Na^+ and similar sized K^+, which are both naturally abundant. Obviously, any conductance of Na^+ and K^+ across AmtB is prevented by the hydrophobic pore interior, where AmtB utilizes the titrability of NH_4^+ and conducts Amm as NH_3. However, how and where AmtB differentiates between the substrates is not yet fully understood. Based on electrostatic calculations it has been suggested that the hydrophobic constriction functions as a filter selective against Na^+ ions while selectivity against K^+ could not be explained [18]. On another hand, calculated relative free energy differences $\Delta\Delta G$ for binding of NH_4^+ vs. monovalent alkali cations at the Am1 site have shown that all the cations have some affinity for this site though NH_4^+ is the favored species [74]. Inhibition of AmtB uptake activity of MeAmm by NH_4^+ [29, 34], which also is observed in experiments on other AMT proteins [4, 26, 27, 172, 173], indicates that Amm is a better AmtB substrate than MeAmm. Moreover, in the experiments on other AMT proteins only small or no inhibiton by alkali metal cations is observed [4, 27, 172, 173], and therefore the Am1 site in AmtB can also be expected to be highly NH_4^+ specific.

The focus of the present study is to quantify the relative importance of intermolecular interactions (among them cation-π interactions) with respect to substrate binding, to elucidate how AmtB differentiates between substrates. For that purpose we resort to Car-Parrinello MD (CPMD) simulations, which in contrast to classical MD simulations, is an *ab initio* MD method and inherently allows polarization of atoms, e.g., polarization of aromatic rings by cations (cation-π interactions), as well as the breaking and formation of covalent bonds, e.g., proton transfer. The latter allow us to address possible proton transfer mechanisms. The limitation of all MD methods is the accessible time scale. Here, the CPMD simulations will be limited to lengths of a few picoseconds and hence should be viewed as a means of "dynamical" optimization of the chosen starting structure rather than a sampling of distinct configurations. We conducted CPMD simulations with four different substrates; NH_3, NH_4^+, Na^+, and K^+, positioned either at the Am1 site or at the Am1b site, as well as NH_4^+ positioned at the Am1a site.

6.2 Materials and Methods

This section is divided into subsections covering details of modeling, simulation, calculation of interaction energies and coordination numbers.

6.2.1 Modeling

Classical molecular dynamics simulations of the *E.coli* AmtB trimer (pdb 1U7G [17]) in a fully hydrated lipid bilayer have been reported elsewhere (see Chapter 4) [75]. Single monomers with NH_4^+ positioned either at the periplasmic binding site (Am1) or at an intraluminal position (Am1a or Am1b) provided equilibrated starting structures for the present CPMD simulations (Fig. 6.1A–B). Following the procedure of Jensen et al. an orthorhombic simulation box was constructed by stripping off the lipid bilayer and only including water molecules within the minimum and maximum x,y coordinates of the monomer (Fig. 6.1A) [174]. Overall neutrality was ensured by replacing water molecules with chloride ions. Geometry optimized structures of NH_4^+ at Am1 and Am1b positions served as template structures for introducing NH_3, Na^+ and K^+ at these sites. At the Am1a position only NH_4^+ was considered. The simulations will be referred to as the Am1, Am1a, and Am1b simulations, respectively.

Equilibrated Starting Structures vs. Crystal Structures

In the equilibrated starting structures used for the Am1 simulations, there are a few differences with respect to the crystal structures of AmtB [17, 18]. A162:C=O is pointing into the substrate binding site and not in-between the F107/F215 stack (Fig. 6.1B–C, see also Chapter 4) [75]. The hydrogen bond between W148:H_ϵ and D160:O is lost due to the intrusion of a water molecule. This leads to a small displacement of W148 and it allows the formation of a two-water wire (w4–w5) connecting NH_4^+ at the Am1 position to the D160 carboxylate group (Fig. 6.2A), as also observed by Lin et al. [76].

In the structure used for the Am1a and Am1b simulations pore blocking F107 and F215 are rotated into a more open state and consequently A162:C=O points directly into the pore lumen instead of in-between the F107/F215-stack (Fig. 6.1D). Additionally, D160 is more solvent exposed and the distance between a D160 carboxylate oxygen and G163:H is shortened [75].

6.2.2 Simulations

Combined first principles MD and classical mechanical MD simulations were conducted using the quantum mechanical (QM) and molecular mechanical (MM) hybrid method as implemented in the Car-Parrinello MD program

(CPMD) [109–111]. Hybrid QM/MM methods allow one fragment of a system to be treated by QM methods, while the rest of the system can be treated by computational less expensive MM methods [110].

Choosing the QM fragment

Identical QM fragments were chosen for all four substrates when the substrates were at identical positions in the channel (Fig. 6.1C–D). Besides the substrate itself, residues (amino acids + water molecules) directly coordinating to the substrate or within a few Ångströms thereof, i.e., potentially able to coordinate to the substrate, were included in the QM fragment. Included were also residues previously proposed to be important for stabilizing the substrates (see Table 6.1), but within a distance of 5 Å with respect to NH_4^+ in the template structures. One exception is F103 at the Am1 site, which was excluded since previous *ab initio* calculations showed that F103 did not contribute significantly in the stabilization of substrates [73].

Substrate position	**QM fragment**			
	Residue	Segment	Link atom(s)	Assigned charge
Am1	$[NH_3/NH_4^+/Na^+/K^+]$	-	-	[0/+1/+1/+1]
	w1–w4 ($4\times H_2O$)	-	-	
	Q104	side chain	C_γ	
	F107	side chain	C_α	
	W148	side chain	C_β	
	S219	side chain	C_α	
	A162-G163	peptide bond	C_α-C_α	
Am1a	$[NH_4^+]$			[0]
	otherwise as Am1b below			
Am1b	$[NH_3/NH_4^+/Na^+/K^+]$	-	-	[-1/0/0/0]
	$2\times NH_3$	-	-	
	D160	side chain	C_β	
	H168	side chain	C_α	
	W212	side chain	C_α	
	F215	side chain	C_α	
	N216	side chain	C_α	
	T273	side chain	C_α	
	H318	side chain	C_α	
	A162-G163	peptide bond	C_α-C_α	

Table 6.2: QM fragments. The residue segments constituting the overall QM fragment, the atoms covalently linking the QM and MM fragments, and the overall charge assigned to a so-called super cell encompassing the QM fragment in the simulation, are listed for each substrate position. Substrates considered at each position and corresponding super cell charges are surrounded by brackets.

The distant D160 carboxylate group was included in the Am1a and Am1b simulations to allow proton transfer via the A162-G163 peptide bond [75], but not in the Am1 simulations since water is a better proton acceptor at that position (see section 6.1). Moreover, in the Am1a and Am1b simulations, H318 and a non-coordinating intraluminal NH_3 molecule were included to avoid instability encountered when treating them as MM atoms. The resulting QM fragments given each substrate position are summarized in Table 6.2.

Computational Details

The QM fragment was treated by *ab initio* density functional theory based on plane-wave expansions and the electron density of the ground state was propagated according to the Car-Parrinello scheme [88, 101, 111]. Utilizing the Born-Oppenheimer approximation, the dynamics of the atoms in the QM fragment is then governed by the instantaneous electronic potential. The dynamics of atoms in the MM fragment was governed by the classical mechanical force field implemented in GROMOS96 [104], employing the AMBER PARM99 parameter set distributed with AMBER 7.0 [175–177]. The simulations were performed within a fully Hamiltonian-coupling scheme between the QM fragment and the MM fragment [110], with the QM-MM interface modeled by the use of a monovalent pseudopotential that saturates the QM region [178]. Previous work has demonstrated the reliability of this method in the description of structural, energetic, and dynamic properties of systems of biological interest, including protein channels (see for instance [174, 179–183]).

The QM fragment was treated at the level of Kohn-Sham density functional theory [184]. For the electronic exchange and correlation part of the Hamiltonian the generalized gradient-corrected approximation, following the prescription of Perdew, Burke and Ernzerhoff (PBE) [185, 186], was applied using a 5.0×10^{-4} density cutoff for the calculation of the gradient correction. We employed Troullier-Martins norm-conserving pseudopotentials for the core electrons + nuclei [187]. Kohn-Sham orbitals were expanded in a plane wave basis set with a kinetic energy cutoff of 70Ry, except for Na^+ or K^+ which were modeled with semi-core pseudopotentials requiring a 80Ry cutoff. For carbon, nitrogen and oxygen the width of the ionic charge distribution was set to 1.0, while 1.2 was used for hydrogen and the carbon atoms linking the QM to the MM fragment. The Kleinman-Bylander separation method was used for the calculation of the nonlocal parts of the pseudopotential [188]. Hydrogen atoms were assigned the mass of deuterium to help prevent an undesired coupling between valence electrons and the core electrons + nuclei. Martyna and Tuckerman's method was used for solving the Poisson equation for the QM fragment [189]. In these calculations the fragment was decoupled from its electrostatic images by treating it as an isolated

system within an orthorhombic super cell. The super cell was assigned a net charge corresponding to the overall charge of the QM fragment (Table 6.2). Striving at zero charge density at the cell borders, at least 5 a.u. separation between the QM fragment and the cell borders was ensured.

In the coupling region between the QM and MM fragments covalent bonds were described by an optimized link-atom pseudopotential [190]. The electrostatic interactions between the two fragments were evaluated as follows: *i*) a modified Coulomb interaction between the QM electronic density and the MM atoms within 6 Å from any QM atom. *ii*) Coulomb interactions between point charges centered on the QM atoms and RESP charges on the MM atoms between 6 Å and 8 Å from any QM atom. Electrostatic interactions between a multipole expansion representing the charge distribution of the QM region and the MM atoms beyond 16 Å from any QM atom. Bonded and van der Waals interactions between the QM and MM parts, as well as within the MM region, were accounted for by the AMBER force field. The atom lists governing the electrostatic interactions between the two fragments were updated every 100 steps. Within the MM fragment long range electrostatic interactions were handled by the particle-particle particle-mesh (PPPM) method implemented in GROMOS96. A grid spacing less than 1 Å was ensured by using a $96 \times 96 \times 96$ mesh. Bonded and van der Waals interactions between the QM and the MM fragments were accounted for by the classical force field.

Geometry optimizations of all structures were performed by means of simulated annealing after a preliminary wavefunction optimization. However, changing NH_4^+ to NH_3 by hydrogen deletion hindered convergence of the wavefunction. Therefore, the NH_3 structures were subjected to a few classical MD steps to relax any critical interactions and subsequently the wavefunction could be optimized successfully.

The geometry optimized structures were assigned velocities according to a Maxwell-Boltzmann distribution at 525K. After 500 MD steps the temperature had converged in the range 270-300K. Additional 500 steps of temperature control lifted the temperature to the target value 300K. The subsequently conducted MD simulations were run for at least 2 ps, except the Am1a simulation which was stopped after ∼1.7 ps, since NH_4^+ had moved to the Am1b position (Fig. A5–A8 in App. A). The final structures of the simulations were geometry optimized as well.

The simulations were conducted at constant energy and volume (NVE) conditions, using an integration time step of 3 a.u. (∼0.0241 fs) and a fictitious electron mass of 500 a.u.. With this setup, the total energy and the fictitious kinetic energy of the electrons in all simulations were conserved within 6.8×10^{-6} a.u.·ps^{-1}·atom^{-1} and 1.5×10^{-6} a.u.·ps^{-1}·atom^{-1}, respectively.

6.2.3 Calculation of Interaction Energies

The interaction energy ΔE_{int} between a specific substrate and selected coordinating residues were obtained by means of single point gas phase DFT calculations. Coordinates were taken from geometry optimized structures and ΔE_{int} was then obtained as:

$$\Delta E_{int} = E_{cluster} - (E_{cavity} + E_{substrate}), \tag{6.1}$$

where $E_{cluster}$ is the energy of the full system (substrate + selected residues, hereafter named as "cluster"), and E_{cavity} and $E_{substrate}$ are the energies of the cluster without the substrate and of the substrate by itself, respectively. The pair interaction energy $E_{int}(R)$ between the substrate and a specific residue R, or more precisely the contribution of this specific interaction to the overall interaction energy $\Delta E_{int}(ALL)$, was obtained as:

$$E_{int}(R) \equiv \Delta\Delta E_{int}(R) = \Delta E_{int}(ALL) - \Delta E_{int}(ALL\text{-}R) \tag{6.2}$$

where $\Delta E_{int}(ALL\text{-}R)$ is the interaction energy between the substrate and the diminished cluster devoid of that particular residue. Since we are interested in relative differences in energies, the structures of the isolated fragments were not relaxed before calculating E_{cavity} and $E_{substrate}$.

Interaction energy calculations were performed using the Am1 starting structures and the Am1 final structures (Fig. 6.2A). All QM treated amino acids were included in the clusters together with water molecules in the first solvation shell (see below). Additionally, we included D160 and w5; the second water in the water-wire connecting NH_4^+ to the D160 carboxylate group, since D160 as the only charged residue close to the Am1 site is expected to provide a substantial electrostatic stabilization [74]. Interaction energy calculations were also performed using the Am1a and Am1b starting structures, and the Am1b final structures (as discussed above, the Am1a final structure was almost identical to the Am1b structures). When referring to an interaction energy as being *smaller* or *larger* than some value, it will be implicit that it is an attractive interaction $E_{int} < 0$, which is numerically *smaller* or *larger* than the given value.

6.2.4 Coordination Number Calculations

Coordination numbers were calculated by applying cutoff distances, which encompassed the first solvation shell. The distances were determined as first minimum in radial pair correlation functions obtained from previous CPMD simulations of the charged substrates in bulk water [191, 192]. Cutoff distances and the corresponding bulk water coordination numbers of the first solvation shell are listed in Table 6.3. In bulk water, four water molecules H-bond to NH_4^+ forming a rigid tetrahedral cage, while a fifth more mobile

water is positioned at the tetrahedral faces [192]. This results in a coordination number for NH_4^+ of about 5.3 (cutoff distance d(N–O)∼3.25 Å) or 4.0 (cutoff distance d($H_{NH_4^+}$–O)∼2.5 Å) [192]. The first solvation shell of Na^+ forms a distorted trigonal bipyramidal structure. The rigidity of this structure leads to a well defined cutoff distance of d(Na–O)=3.1 Å and corresponding coordination number of 5.2 [191]. The first solvation shell of K^+ lacks such geometry and is more floppy, which yields a less well defined cutoff distance [191]. Consequently, the coordination number will be more sensitive to the applied cutoff distance [191]. Here, the cutoff distance d(K–O)=3.6 Å is chosen, corresponding to a coordination number of ∼6.7 [191]. Similar data were not available for NH_3 and we therefore applied a cutoff distance identical to that of NH_4^+.

For cation-π interactions we applied cutoff distances adopted from gas phase DFT calculations at the B3LYP/6-31G** level of theory reported for complexes between cations and aromatic amino acid motifs (F,W) [193] (Table 6.3).

Position		**Type**	**NH_3**	**NH_4^+**	**Na^+**	**K^+**
Am1	{	protein	0.6±0.5	1.2±0.6	0.2±0.4	1.0±0.0
		prot+wat	2.9±0.8	3.5±0.9	4.3±0.5	5.1±0.5
Am1a		prot+wat	-	3.4±0.5	-	-
Am1b		prot+wat	0.8±0.6	2.6±0.7	3.0±0.0	3.0±0.0
Bulk		water	-	4.0[a]	5.2[b]	6.7[b]
cutoff (Å)		F:π	-	2.9[c]	2.4[c]	2.9[c]
		W(six-ring):π	-	2.9[c]	2.3[c]	2.8[c]
		other	2.5	2.5[a]	3.1[b]	3.6[b]

Table 6.3: Average coordination numbers. Average coordination numbers and standard deviations for the four substrates at the Am1 and Am1b positions and for NH_4^+ initially at the Am1a position, listed together with the bulk hydration numbers (first solvation shell) obtained from other CPMD simulations [191, 192]. "Type" denotes the coordination number to water, protein or both (prot+wat). Additionally, the applied cutoff distances in Å for cation-π interactions and for all "other" interactions are listed. [a][192], [b][191], [c][193].

6.3 Results and Discussion

The results presented mainly focus on the optimized (final) structures derived from the Am1, Am1a, and Am1b simulations. Differences in geometries, coordination patterns, and interaction energies between substrates and coordinating residues will be discussed. We seek to estimate the relative importance of coordinating residues on substrate recognition and recruitment. Proposed proton transfer pathways will also be discussed. Tables listing coordination distances and interaction energies of the optimized initial structures as well as figures showing the geometry of the initial structures are provided in Appendices (Figs. A9–A10 and Tables B1–B3 in App. A and B,

respectively). Coordination distances and coordination numbers along the trajectories of all simulations are provided in Figs. A1–A8 in App. A.

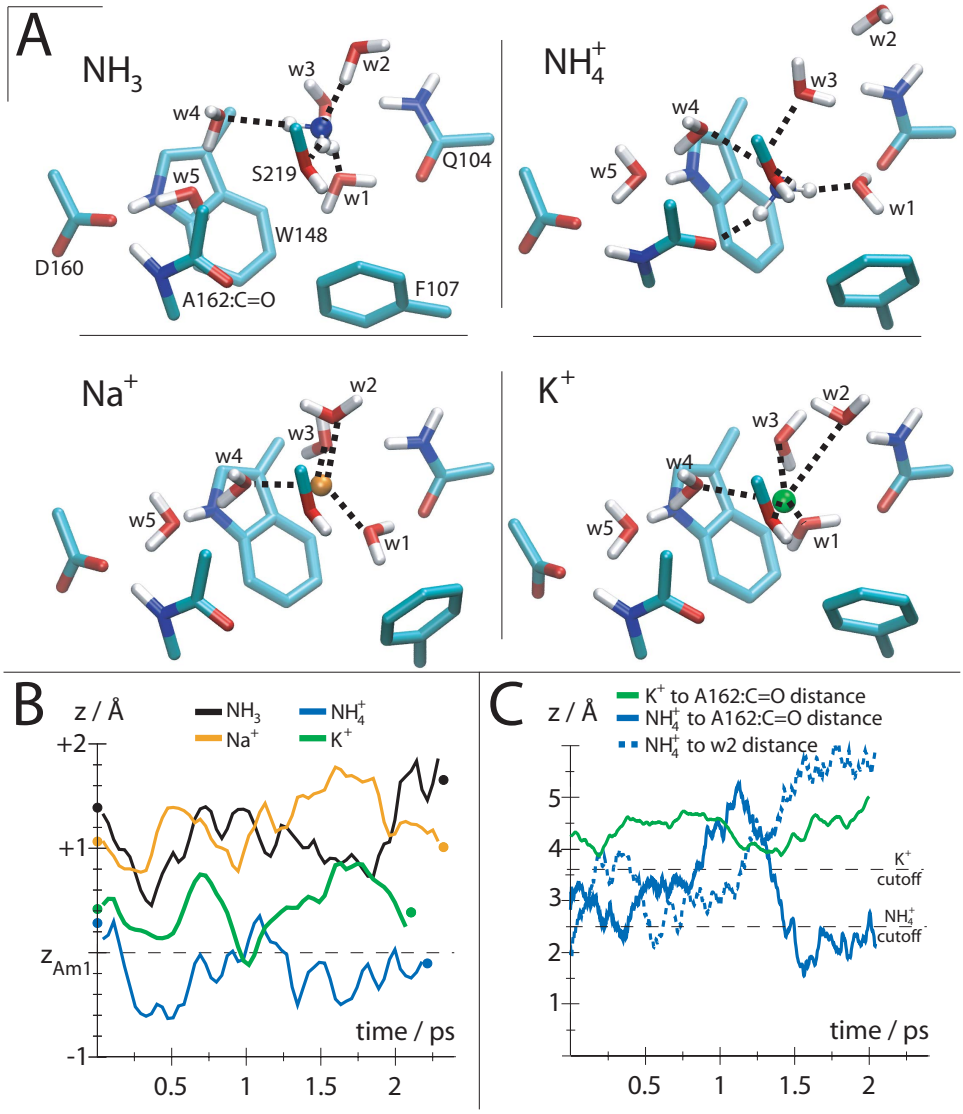

Figure 6.2: Substrates at the Am1 position. (A) Optimized final structures after ∼2 ps simulation with the four different substrates (*ball-and-stick*): NH_3, NH_4^+, Na^+ (*orange*), and K^+ (*green*), initially at the Am1 position. Residues within coordination distances are connected to the substrates by dashed lines. (B) z-positions of substrate heavy atoms relative to the Am1 position along the trajectories and z-positions in the optimized structures (circles). (C) The distances between NH_4^+ and A162:C=O, and water molecule w2, along the trajectory reveal a correlation between the loss of a coordinating water and coordination to A162:C=O. Additionally, the distance between K^+ and A162:C=O is shown.

6.3.1 The Am1 Site Selects NH_4^+ over Na^+, K^+ and NH_3

Figure 6.2B shows the time evolution of substrate positions along the channel axis (z-axis) for the four substrates considered (NH_3, NH_4^+, Na^+ and K^+). It reveals that NH_4^+ and K^+ are the only substrates which reach the Am1 position during the 2 ps simulation. On average NH_4^+ is buried ∼0.2 Å below the Am1 position, while K^+ stays 0.5 Å above. NH_3 and Na^+ both stay ∼1 Å above the Am1 position. This time evolution of substrate positions indicates that NH_4^+ has the highest affinity for the site, followed by K^+, Na^+ and NH_3. That NH_4^+ is the charged substrate with the highest affinity is also indicated by its higher coordination number relative to bulk water, although the standard deviation is somewhat larger than for Na^+ and K^+ (Table 6.3). On average NH_4^+ is better coordinated in the second half of the trajectory, where the average coordination number 3.9±0.7 is very close to the bulk value (4.0) (see, e.g., Fig. A4 in App. A). This is also reflected by the final fivefold coordination of NH_4^+ (Fig. 6.2A). In contrast do the corresponding coordination numbers of Na^+ and K^+ not change significantly when only considering the second half of the trajectories (see, e.g., Fig. A4 in App. A) or the final structures (Fig. 6.2A).

The total interaction energies (see section 6.2) of the optimized initial and final structures do not differ significantly for NH_3 (0.2 kcal/mol) or Na^+ (0.2 kcal/mol), while NH_4^+ (-8.1 kcal/mol) gains ∼8 kcal/mol and K^+ (4.4 kcal/mol) looses 4.4 kcal/mol (Table 6.4, see also Table B3 in App. B). Since it is energetically favorable for NH_4^+ to bury itself below the Am1 position, it supports the above indications that NH_4^+ has a large affinity for the Am1 site. This is further emphasized by the fact that it gains the energy even though a water molecule is stripped off from the first solvation shell, i.e., NH_4^+ only coordinates water by two of its hydrogen atoms instead of three as in the initial structure (Fig. 6.2A,C, water w2). The energy loss of K^+ occurs even though it is located practically at the same position in both the initial and final structure (Fig. 6.2B). Apparently, this is due to an increase in the distance from 2.74 to 3.22 Å to coordinating water w2. Since the first solvation shell of K^+ is less structured than that of NH_4^+ and Na^+, the number of water molecules to be included in the interaction energy calculations and thereby the total interaction energy will be more sensitive to the applied cutoff distance [191, 192]. From visual inspection of the structures, it could be observed that another water molecule had approached but not crossed the cutoff distance in the final structure, which can partly explain the energy loss of K^+.

Position	Fragment (R)	NH_3	NH_4^+	Na^+	K^+
Am1	Q104	-2.2	-8.4	-1.8	-4.3
	F107	0.7	-3.3	-4.1	-2.6
	W148	-0.2	0.9	-4.6	-2.8
	D160	-2.7	-49.9	-46.7	-46.2
	A162:C=O	0.1	-16.9	-1.8	-5.1
	S219	-2.1	-10.1	-5.3	-10.3
	w1	-3.5	-16.5	-22.0	-15.3
	w2	-12.0	NI	-13.7	-4.2
	w3	NI	-4.7	-12.9	-4.4
	w4	-1.6	-9.3	-19.6	-14.0
	w5	-1.1	-6.8	-9.1	-7.7
	ΔE_{int}(ALL)	-21.3	-144.1	-153.8	-128.0
	ΔG_{bind} ($\approx$)	-13	-55	-46	-40
(Am1a) Am1b	A162:C=O	-0.4	(-13.9) -11.2	-25.0	-16.4
	H168	-2.9	(-10.7) -24.1	-26.5	-18.7
	W212	0.0	(5.5) -5.1	-2.1	-2.7
	F215	-1.1	(-2.8) -4.6	-3.8	-6.0
	N216	-0.1	(-13.4) -4.5	-4.8	-5.4
	T273	0.2	(-7.0) -2.1	-1.1	-1.6
	NH_3	0.1	(-22.4) -16.6	-21.2	-13.4
	ΔE_{int}(ALL)	-5.3	(-91.7) -93.4	-105.7	-84.2

Table 6.4: Interaction Energies. Interaction energies E_{int}(R) between substrates and coordinating residues obtained from cluster calculations on the final structures. All energies are in kcal/mol. ΔE_{int}(ALL) is the total interaction energy. Since the energies are not pairwise additive, they do not sum up to a total of ΔE_{int}(ALL), but to $\sim$90% of ΔE_{int}(ALL) for the charged substrates and $\sim$120% of ΔE_{int}(ALL) for NH_3. Also listed is the estimated binding free energy ΔG_{bind} (see text). Entries in parenthesis are the energies of the initial structure of NH_4^+ at the Am1a position. "NI" denote that the residue was not included in the calculation..

	NH_3	NH_4^+	Na^+	K^+
ΔG_{sol}	-4.3[a]	-82.0[b]	-100.1[b]	-82.5[b]
ΔH_{sol}	-7.9[a]	-88.8[b]	-107.8[b]	-87.6[b]
$\Delta H[M(H_2O)_{0\rightarrow n}]$	-	(n=5) -67.3[c]	(n=5) -85.7[d]	(n=6) -79.7[d]

Table 6.5: Thermodynamic Data for Solvation and Hydration of Substrates. Free energies ΔG_{sol} and enthalpies ΔH_{sol} of solvation are listed for all four substrates (M). Hydration enthalpies $\Delta H_{hyd}[M(H_2O)_n]$ associated with formation of the first solvation shell composed of n water molecules are listed for the three charged substrates. [a][194], [b][195], [c][196], [d][197].

What is the reason why AmtB seemingly binds NH_4^+ more strongly than K^+, Na^+ and and NH_3? According to the solvation free energies (Table 6.5, ΔG_{sol}), dehydration is more costly for Na^+ than for K^+ and NH_4^+, and consequently demands a larger compensation from substrate-protein interactions. Of all charged substrates such a compensation seems most inferior for Na^+ (Table 6.3). To rank the substrates in accordance with binding affinities prerequisites knowledge about the binding free energies ΔG_{bind}, which can be calculated as $\Delta G_{bind} = \Delta G_{com} - \Delta G_{sol}$, where ΔG_{com} is the complexation free energy associated with bringing the substrate from the gas phase and into the solvated complex (substrate+protein), here into the Am1 site of AmtB. According to Table 6.5 the entropic contribution $T\Delta S_{sol}$ to $\Delta G_{sol} = \Delta H_{sol} - T\Delta S_{sol}$ is small compared to ΔH_{sol} and approximately of the same size (-5.1 to -7.7 kcal/mol) for all charged substrates. Consequently, we assume that the entropic contributions to ΔG_{com} and therefore ΔG_{bind} can be expected to be relatively small and of the same size for all charged substrates as well. Thus, to obtain a gross estimate of ΔG_{bind} we apply the approximation $\Delta G_{bind} \approx \Delta H_{com} - \Delta H_{sol}$ and use for ΔH_{com} the calculated interaction energy ΔE_{int}(ALL). The results (Table 6.4) show that NH_4^+ has the largest affinity, followed by Na^+ and K^+. Additionally, the contribution from substrate-protein interactions to the coordination numbers relative to the bulk values is largest for NH_4^+ (Table 6.3). When combined, the above results indicate that NH_4^+ is preferred over monovalent cations such as Na^+ and K^+ due to a larger compensation of the dehydration cost from substrate-protein interactions. Seemingly, NH_3 also has a low coordination number with respect to substrate-protein interactions (Table 6.3), and its estimated binding free energy is much smaller than for the charged substrates (Table 6.4). This implies that differentiation between charged and neutral species is due to relatively larger stabilization of charged substrates than of neutral substrates at the Am1 site. That is, if NH_3 would enter the Am1 site, it is likely to be protonated, since NH_4^+ has a larger affinity for the site.

In conclusion, AmtB already at the Am1 site selects NH_4^+ over neutral molecules like NH_3 and over monovalent ions such as Na^+ and similar sized K^+, by means of a fine tuned balance between charge stabilization and dehydration cost. This might explain why monovalent cations do not inhibit MeAmm conduction in experiments on other AMT proteins [4, 27, 172, 173].

6.3.2 D160 Stabilizes Charged Species in The Periplasmic Vestibule

Table 6.4 lists the interaction energies between the individual substrates and coordinating residues at the Am1 site. These energies are correlated with distances in Fig. 6.3. Even though D160 is distant (>6 Å) from the substrates its carboxylate group provides the largest interaction energy for any

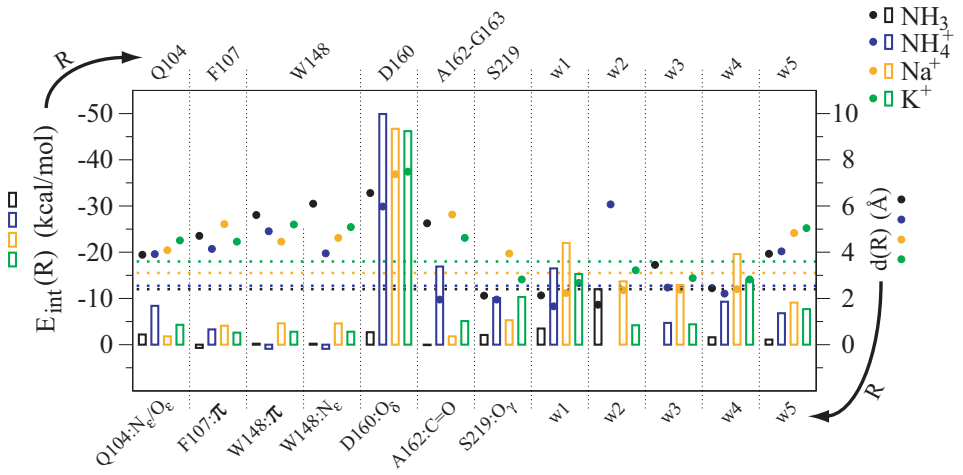

Figure 6.3: Interaction energies and distances at the Am1 site. For all four substrates interaction energies $E_{int}(R)$ (*boxes*) and distances d(R) (*circles*) to residues R are plotted. Dashed lines are individual cutoff distances applied for the calculation of coordination numbers. Interaction energies are listed in Table 4, while distances are listed in Table B1 in App. B.

of the charged substrates, i.e., it contributes with -45 to -50 kcal/mol corresponding to ∼30–40% of the total interaction energy or more than twice the size of any other interaction energy. Hence, it is the single most important interaction for stabilizing positively charged substrates in the binding pocket, which is in accordance with related observations [74, 77]. Interactions to Q104, F107, W148, A162-G163, and S219 also provide some stabilization of the charged substrates. The importance of cation-π interactions involving F107 and W148 will be discussed later.

6.3.3 NH_4^+ Recognition - A162:C=O Plays an Important Role

In all Am1 simulations, A162:C=O rotates toward its crystal structure conformation pointing in-between the F107/F215 stack (Fig. 6.2A) and, correspondingly, the interaction energies to the substrates decrease (see, e.g., Table B3 in App. B). Only in the NH_4^+ simulation A162:C=O rotates back again and a good hydrogen bond to NH_4^+ is formed in the final structure (Fig. 6.3). This is associated with a favorable increase in interaction energy by ∼10 kcal/mol (see, e.g., Table B3 in App. B). Due to the improved interaction with A162:C=O, NH_4^+ moves below the Am1 position and concurrently strips off water w2 from its first solvation shell (Fig. 6.2B–C), as also seen elsewhere [76]. Hence, A162:C=O provides the main driving force for the dehydration, i.e., provides the main compensation of the dehydration cost, which was considered important for the selection of NH_4^+ above.

The ability of A162:C=O to coordinate NH_4^+ at the Am1 position, as well

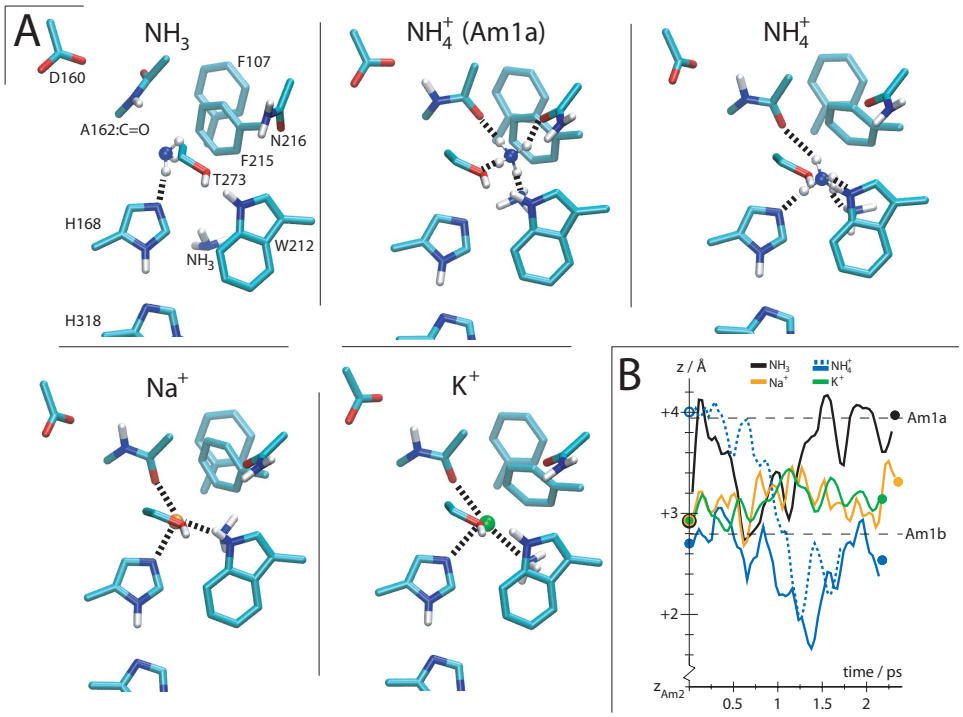

Figure 6.4: Substrates at the Am1a and Am1b positions. (A) Optimized final structures after ∼2 ps simulation with the four different substrates(*ball-and-stick*): NH_3, NH_4^+, Na^+ (*orange*), and K^+ (*green*), initially at the Am1b position. Only the optimized initial structure of NH_4^+ at the Am1a position is shown since it translocated to the Am1b position during simulation. Residues within coordination distances are connected to the substrates by dashed lines. Interaction energies between substrates and residues shown are listed in Table 6.4. Interaction energies are correlated with distances in Fig. 6.5, while distances are listed in Table S2 (Suppl. Mat.). The intraluminal NH_3 coordinating to the substrate is shown in licorice. (B) z-positions of substrate heavy atoms relative to the Am2 position along the trajectories as well as the z-positions in the optimized Am1a (empty circles) and Am1b structures (filled circles).

as to reorient from intra- to extraluminal positions, supports its important role in NH_4^+ stabilization [74–76], and in guiding NH_4^+ across the F107/F215 stack (see Chapter 4 and 5) [75]. This is further verified, by a comparison of the optimized initial and final structures of the Am1a and Am1b simulations, where A162:C=O initially is directed into the pore lumen (Fig. 6.1D); the conformation is stable for all charged substrates, but not for neutral NH_3 where A162:C=O reorients toward the Am1 site (Fig. 6.4A).

A162:C=O does not come within the cutoff distance for the first solvation shell of Na^+ nor of K^+, even though K^+ is positioned at or close to the Am1 site during simulation (Fig. 6.2C, see also Fig. A1 in App. A). Seemingly the nature of the Am1 site only allows cations capable of donating hydrogens to interact strongly with A162:C=O. This provides a means of recognizing NH_4^+, as discussed above, from monovalent alkali metal cations

before translocation of substrate into the channel lumen.

6.3.4 Q104 Stabilizes the Substrate

Comparison of structures, distances and corresponding interaction energies show that Q104:O_ϵ has some effect in polarizing w1 and thereby in improving the interaction between w1 and the substrates. In fact, for all charged substrates w1 provides the largest interaction energy of all coordinating water molecules (Table 6.4 and Fig. 6.3, w1). In decreasing order, the polarization effect is most important for the final structures of NH_4^+, K^+ and least for Na^+ (Table 6.4 and Fig. 6.3, Q104). Both NH_4^+ and K^+ are close to the Am1 position and NH_4^+ has lost water w2 from the first solvation shell as previously discussed, while for K^+ the distance to water w2 is increased. Hence, the polarization effect seems to become more important as the substrate comes close to the Am1 position and as it dehydrates.

Thus, in addition to its possible role in substrate differentiation by steric exclusion effects (see Chapter 4) [75], Q104 may also indirectly play a role in stabilizing NH_4^+ along the first part of the translocation pathway by polarizing a coordinating water molecule. The Q104 conformation resembles the conformation encountered in the crystal structures of wild type AmtB [18] and not of engineered AmtB [17]. Upon superposition of the crystal structures, it is obvious that the wild type conformation of Q104 is sterically possible only when NH_4^+ and not $MeNH_3^+$ occupies the Am1 site. Hence, substrate stabilization due to water polarization by Q104 can be expected to have less effect for $MeNH_3^+$ than for NH_4^+ though polarization and steric exclusion effects are not mutually exclusive. This provides partly an explanation to the fact that NH_4^+ is a better AmtB substrate than $MeNH_3^+$ [29, 34].

6.3.5 Are Cation-π Interactions Important?

None of the charged substrates come within cation-π interaction distances to aromatic residue F107 and W148 in the Am1 simulations, nor to F215 and W212 in the Am1a and Am1b simulations (Figs. 6.3 and 6.5, see also Figs. A3 and A7 in App. A–B). For all charged substrates at the Am1 site the magnitude of the calculated interaction energies to F107 or W148 are smaller than 4.6 kcal/mol and when combined, substrate interactions with these residues account for at most 6 % of the total interaction energies (Table 6.4). Correspondingly, at the Am1b site the magnitude of the interaction energies to F215 or W212 are smaller than -6.0 kcal/mol and when combined account for at most 7% of the total interaction energies (Table 6.4). Compared with the other interaction energies at the Am1 site (Table 6.4 and Fig. 6.3) and at the Am1a and Am1b sites (Table 6.4 and Fig. 6.5), cation-π interactions are clearly not among the most important interactions. Our results for the Am1

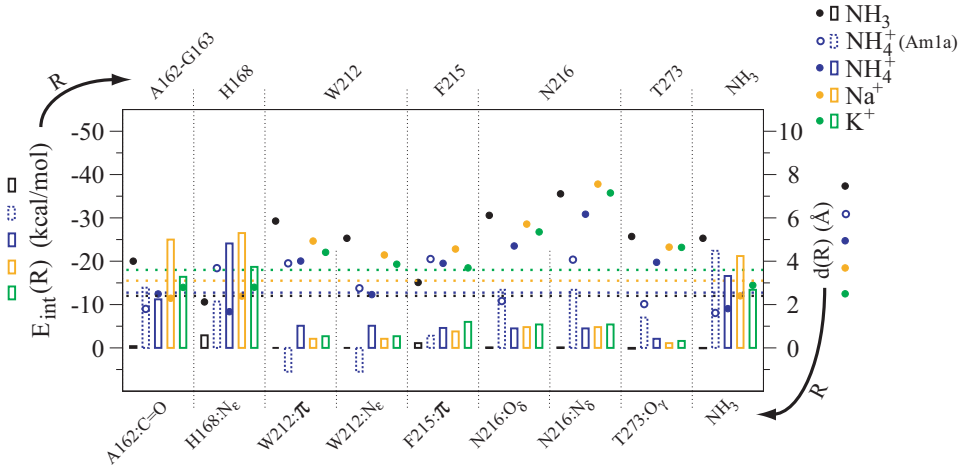

Figure 6.5: Interaction energies and distances at the Am1a and Am1b sites. For all four substrates interaction energies $E_{int}(R)$ (*boxes*) and distances d(R) (*circles*) to residues R are plotted. Dashed boxes and empty circles represent data for NH_4^+ at the Am1a position. Dashed lines are the individual cutoff distances applied for the calculation of coordination numbers. Interaction energies are listed in Table 4, while distances are listed in Table B2 in App. B.

site contrast with a previous *ab initio* study, which concluded that cation-π interactions provide a large stabilization of substrates at this position [73]. The causes of these discrepancies are discussed below.

In the present study, the interaction energies determined for F107 and W148 in the initial and final structures are -2.0 (F) and -3.6 (W) kcal/mol, and, -3.3 (F) and 0.9 (W) kcal/mol, respectively (Table 6.4 and Table B3 in App. B). These energies differ considerably from the values -10.9 (F) kcal/mol and -17.4 (W) kcal/mol reported by Liu and Hu [73]. In the crystal structure, the shortest distance between the NH_4^+ nitrogen and W148 is 3.63 Å to the six-membered ring, while in our optimized initial and final structures the distance is 4.67 and 4.91 Å to the five-membered ring, respectively (Fig. 6.3, see also Table B1 in App. B). Therefore, the small deviation of the optimized structures from the crystal structure could be the reason for this discrepancy. Such an explanation does not apply for F107, where the crystal structure distance (3.97 Å) is comparable to the initial (4.29 Å) and final (4.14 Å) structure distances, and consequently similar interaction energies are expected (Fig. 6.3, see also Table B1 in App. B). Additionally, NH_4^+ and K^+ in the Am1b simulations adopt ring centered positions relative to F215 with even shorter distances but the corresponding interaction energies of -4.6 kcal/mol and -6.0 kcal/mol, respectively, are still considerably smaller than those of Liu and Hu [73] (Fig. 6.5 and Table 6.4). This rules out structural differences as the cause of the observed discrepancies, which we think rather result from different methods used to calculate interaction

energies. Liu and Hu calculate the interaction energy between the substrate A and a coordinating residue B as E_{int}=E_{AB}-(E_A+E_B) [73]. Using this approach, polarization effects from surrounding residues are not explicitly taken into account, which results in large interaction energies. Furthermore, only F103, F107, W148 and S219 were included in their calculations, which thereby neglect the important contributions from Q104, A162:C=O and first most D160.

In conclusion, we find that cation-π interactions provide stabilization of charged substrates at the Am1 site, but in contrast to previous suggestions they do *not* provide the main contribution. It is likely that the cation-π interaction to F107 gradually becomes more important as NH_4^+ translocates below the Am1 position; F107 will then need to rotate and NH_4^+ may in turn gain a more optimal position for cation-π interactions with F107 and, eventually, F215.

6.3.6 Intraluminal NH_3 is Gaseous

According to the small total interaction energy of NH_3 (Table 6.4) and large coordination distances (Fig. 6.5), NH_3 is weakly bonded already in the upper part of the pore lumen (Fig. 6.4A–B). This is also reflected by its low coordination number relative to that at the Am1 position (Table 6.3). Only H168 and probably other intraluminal NH_3 molecules provide some stabilization (Table 6.4). This complies with the view that NH_3 is in a gaseous state when occupying the pore lumen [17].

6.3.7 Charged Substrates at Intraluminal Positions!

As argued above, differentiation between NH_4^+ and Na^+, as well as K^+, most likely occur around the Am1 site. However, consideration of intraluminal Na^+ and K^+ can still be useful for investigating a charged substrate, i.e., NH_4^+, positioned intraluminally.

The final structures of the Am1b simulations show that all charged species cluster around the Am1b positions (Fig. 6.4A–B). Both Na^+ and K^+ are on average localized at ∼0.4 Å above, while NH_4^+ on average is positioned ∼0.3 Å below the Am1b position. In the simulation with NH_4^+ initially at the Am1a position, NH_4^+ moves further down in the pore lumen after ∼1 ps (Fig. 6.4B, *dashed line*). Thereby it looses coordination to N216:O_δ and T273:O_γ, but gains one to H168:N_ϵ, rendering coordination patterns and positions similar to those of NH_4^+ initially at Am1b.

The average coordination numbers of the Am1b simulations (Table 6.3) and the final coordination (Fig. 6.4A) reveal that NH_4^+ is again the best accommodated charged substrate relative to the bulk value. In fact the table shows that the coordination number calculated for the Am1a simulation is comparable to that of the Am1 site (see, e.g., Fig. A4 in App. A). Relative

to the bulk value especially K^+ is poorly coordinated and this opposes that it can be accommodated at intraluminal positions, which could not be ruled out in previous electrostatic calculations [18]. As previously discussed, the differentiation between (charged) substrates most likely occur already before translocation across the F107/F215 stack. Hence, occupation of intraluminal positions by Na^+ and K^+ most likely has little or no physiological relevance.

6.3.8 Proton Transfer

Proton transfer from NH_4^+ to surrounding amino acids or water molecules was not observed in any of the simulations performed. Therefore, we manually moved the proton to D160 in the Am1a structure to see if the resulting proton configuration was energetically stable. The proton transferred back to NH_3 immediately, suggesting that NH_4^+ is favored over NH_3 at the Am1a and Am1b positions and that proton transfer presumably must take place above the Am1a site. Alternatively, since H168:N_ϵ provides the largest stabilization of charged species around the Am1b position (Table 6.4 and Fig. 6.5), a different neutral protonation state of intraluminal Histidines, i.e., H168:N_ϵH-H318:N_δ, is ought to favor NH_3 over NH_4^+ at this position and may drive the deprotonation of an intraluminally NH_4^+ positioned around Am1a. Indeed, previous pK_a shifts calculation have shown that the preferred neutral protonation states of intraluminal Histidines are sensitive to the presence/absence of intraluminal ammonia and water (see Chapter 4) [75].

Since we did not include D160 and w5 in the QM fragment of the Am1 simulations, we cannot exclude that a proton can be transferred to D160 via a two-water wire [76]. However, it is unlikely that NH_4^+ deprotonates along that pathway (see section 6.1). Regarding the catalytic function of S219 [77], no proton transfer to S219:O_γ occurred in the Am1 simulation. In fact, the distance to coordinating water w1 was shorter throughout the simulation (Figs. A1–A2 in App. A). Hence our simulations do not support the suggested catalytic function of S219 with respect to proton transfer [77], at least not for the positions and conformations adopted by NH_4^+ and S219 in the present study, respectively. Whether the proton then will transfer from NH_4^+ directly to periplasmic water [80] or indirectly via the A162-G163 peptide bond to D160 or eventually to a periplasmic water molecule [75] remains an open question. The critical point is whether NH_4^+ is hydrated or not before the proton transfer takes place. If NH_4^+ is (partially) hydrated then the proton will transfer directly to water, otherwise it is more likely to transfer via the A163-G163 peptide bond. Since cation-π interactions are not well represented by additive classical mechanical force fields [160] the optimal position for deprotonation of NH_4^+ previously suggested in Refs. [76, 80] and in Chapter 5, is encumbered with uncertainty. Therefore, further *ab initio* MD studies could elucidate at which intermediate position between Am1 and

Am1b NH_4^+ becomes deprotonated, how it relates to the neutral protonation state of H168:N_ϵ, and what the preferred proton transfer pathway is.

6.4 Conclusion

In this work we have conducted Car-Parrinello molecular dynamics simulations of *E.coli* AmtB with the four different substrates NH_3, NH_4^+, Na^+ and K^+ positioned at either the periplasmic substrate binding site Am1 or at intraluminal position Am1b. Additionally, one simulation with NH_4^+ at intraluminal position Am1a was conducted. Our main motivation was to clarify the relative importance of different interactions with respect to substrate binding and selectivity, as well as test some recently proposed proton transfer mechanisms. These are some of the main questions of the biophysical research on this protein.

CPMD simulations were limited to lengths of ∼2 ps and should be viewed as means of a "dynamical" optimization. To better resolve effects of accommodating different substrates at each position the simulations were started from identical system configurations.

For each substrate average coordination numbers along the trajectories were calculated and for each initial and final optimized structure, pair interaction energies as well as total interaction energies between the substrate and surrounding residues were calculated. Differences in geometries, coordination patterns, and interaction energies between substrates and coordinating residues provided the basis for estimating the relative importance of coordinating residues on substrate recognition and recruitment.

We find that D160 contributes the most in stabilizing and recruiting charged substrates at the recruition (Am1) site; it accounts for 30–40% of the total interaction energy between the substrates and surrounding residues. Interestingly, previously proposed cation-π interactions to W148 and F107 only account for ∼6% of the total interaction energy. Thereby, we are able to shed light on the importance of cation-π interactions; these do *not* provide the main contribution for binding charged substrates at the Am1 position. This disagreement with previous results from similar studies can be attributed in part to the fact that the previous calculations did not take polarization effect from surrounding residues into account, nor did these calculations include key residues. We also find that Q104 can provide stabilization of substrates through water polarization. The effect is expected to have less effect for $MeNH_4^+$ than for NH_4^+, and may provide some explanation of the experimentally determined differences in Amm and MeAmm affinity [29, 34].

Of the four different substrates, we find that NH_4^+ has the largest affinity for the Am1 site, while K^+, Na^+ and NH_3 have less affinity. The smaller affinities of K^+ and Na^+ are mainly due to less compensation of their dehy-

dration cost from substrate-protein interactions, while the smaller affinity of NH_3 mainly is due to a larger stabilization of charged substrates relative to neutral substrates at the Am1 site. We find that AmtB differentiates between NH_4^+ and the monovalent alkali metal ions by allowing the carbonyl group of A162 to interact strongly with NH_4^+ but not the other ions. Thus, A162 may select cations that are capable of donating a hydrogen over other cations in agreement with inhibition studies on other AMT proteins [4, 27, 172, 173]. Our findings support that A162 coordinates NH_4^+ at the Am1 site as observed in classical MD studies [18, 74–76]. Likewise, we observe that the coordination to A162 drives a partial dehydration of NH_4^+ [76]. Additionally, our simulations support that A162 can play an active role in translocating NH_4^+ across the F107/F215 stack [75].

At the outer part of the lumen, Na^+ and K^+ are both poorly coordinated, while NH_4^+ is reasonably well coordinated. Hence, if not already deprotonated NH_4^+ may in principle accommodate the Am1a and/or Am1b positions. Intraluminally, we find that NH_3 is gaseous, which conforms to a previous view [17].

Since NH_3 is the net transported species, it is of great interest to know how and where NH_4^+ deprotonates between the Am1 site and the outer part of the lumen. Here we tried to test our previously proposed proton transfer pathway [75], but no immediate support for this pathway was found. However, we cannot rule out that NH_4^+ deprotonates along this pathway at positions above the Am1a position or, alternatively, that a different neutral protonation state of intraluminal Histidines H168 and H318 drives the deprotonation of NH_4^+ around the Am1a and Am1b positions. Neither, did our simulations provide any immediate support for a proposed catalytic role of S219 [77], and they are inconclusive with respect to a proton transfer directly to periplasmic water [80]. Hence, further investigations based on *ab initio* MD simulations are needed to determine the exact position where NH_4^+ deprotonates and the preferred proton transfer pathway.

Chapter 7

Conclusion

This thesis summarizes my work on *Escherichia coli* AmtB; a transmembrane protein belonging to the ammonia/ammonium transporter (AMT) family of proteins. This AMT protein was studied by means of classical and *ab initio* molecular dynamics (MD) simulations in order to investigate the coupling between substrate transport and protein structure dynamics. Concluding remarks have already been provided in the individual conclusions of Chapters 4–6. Here, I will therefore just briefly summarize some of the main findings of my work on AmtB and relate them to the questions asked about AmtB function in the introduction in Chapter 1.

How does AmtB recruit substrate? It has previously been suggested that the aromatic rings of F103, F107 and W148 and the hydrogen bond acceptor S219:O_γ constitute a high affinity periplasmic substrate binding site selective of cations. The simulations showed that cation-π interactions with the aromatic rings do not contribute much in the binding of cations, while S219:O_γ does. However, the most important contribution for stabilizing and recruiting positively charged substrates was found to be an electrostatic interaction to the negatively charged carboxylate group of D160. Additionally, the simulations revealed that A162:C=O is important for binding NH_4^+ and in guiding NH_4^+ across the F107/F215 stack.

How does AmtB recognize substrate? From experiments it is known that the physiologically important monovalent cations Na^+ and K^+ do not inhibit MeAmm transport in other AMT proteins, but that Amm does (also in AmtB) [29, 34]. Although the inhibitory effect of Na^+ and K^+ on MeAmm conduction in AmtB can only be deduced from experiments on other AMT proteins [4, 27, 172, 173], the simulations indicate that the dehydration costs of Na^+ and K^+ are less compensated by substrate-protein interactions than that of NH_4^+, suggesting that also in AmtB Na^+ and K^+ will have small or no inhibitory effects on Amm conduction. A162:C=O seems to play an important role in substrate recognition by selecting cations that are capable

of donating a hydrogen over other cations. The simulations showed that the inhibition by Amm might at least partially be explained by two different effects, which both involve Q104 and both favor NH_4^+: *1*) a stabilization effect due to polarization of a water coordinating to NH_4^+ and *2*) a steric exclusion effect.

Where and how is NH_4^+ deprotonated? Experiments have indicated that protons are not co-transported with NH_3 in AmtB, but must be delivered back to periplasm. In the "equilibrium" classical MD simulations of Chapter 4 two intraluminal binding sites, Am1a and Am1b, providing tetrahedral coordination of NH_4^+ were identified. On that background a proton transfer mechanism was proposed. In this mechanism the proton is exchanged at the A162-G163 peptide bond via an imidic acid mechanism catalyzed by the D160 carboxylate group and intraluminally positioned NH_4^+. Subsequently, both the localization of the deprotonation region and the mechanism were subject to further investigations. The localization of the deprotonation region was estimated from calculated free energy profiles obtained from steered MD simulations. The mechanism was tested by conducting hybrid QM/MM MD simulations. These investigations confirmed that the deprotonation region encompasses an intraluminally positioned NH_4^+, but the mechanism could not be confirmed nor rejected.

The work presented in this thesis have added some pieces to the puzzle of AMT function, but the puzzle is still not complete. Especially, the exact mechanism by which NH_4^+ is deprotonated remains to be established and requires future investigations.

Publications & Manuscripts

This thesis is based on research that has been carried out in the period January 2004 to June 2007, and has resulted in the following publications and manuscripts:

1 Nygaard, T. P., C. Rovira, G. H. Peters, and M. Ø. Jensen. 2006. Ammonium Recruitment and Ammonia Transport by E.coli Ammonia Channel AmtB. *Biophysical Journal* 91:4401–4412.

2 Nygaard, T. P., M. Ø. Jensen and G. H. Peters. The effect of Asp54 phosphorylation on the energetics and dynamics in the response regulator protein Spo0F studied by molecular dynamics. *Submitted*

3 Nygaard, T. P., G. H. Peters, and M. Ø. Jensen, and C. Rovira. Ab Initio Molecular Dynamics Study of Substrate Recognition and Cation-π Interactions in E.coli Ammonia Channel AmtB. *In preparation*

Dansk Resumé

Denne afhandling omhandler molekyldynamiske (MD) simuleringer af *E.coli* Ammoniakkanalen AmtB, som tilhører AMT familien af transmembrane ammoniak/ammonium (Amm) transporterende proteiner. Ammoniak (NH_3), som i dets protonerede form er ammonium (NH_4^+), er en vigtig næringskilde for mikroorganismer og planter. I pattedyr er Amm hovedsagligt et stofskifte-affaldsprodukt, men har dog også nogle vigtige funktioner bl.a. i nyrerne, hvor udskilning af NH_4^+ til urinvejene er med til at regulere syre-base balancen i blodet. Dog er Amm giftigt i forhøjede koncentrationer der som oftest er en følge af nedsatte Amm reguleringsfunktioner i kroppen, f.eks. nedsat nyrefunktion, og menes at være en medvirkende årsag til hjernelidelser så som Alzheimer type II. At Amm-regulering og -transport er fysiologisk vigtige mekanismer afspejles af, at alle levende organismer har proteiner tilhørende AMT familien siddende i deres cellemembraner. I mennesket har det vist sig, at Rhesus proteinerne, som hovedsageligt kendes fra Rhesus-blodtypesystemet og deres medvirken til røde blodlegemers immunreaktion, også tilhører AMT familien. AMT proteiners funktion i Amm-transport over cellemembraner gør dem bl.a. interessante i forbindelse med optimering af nitrogenoptagelse/-udnyttelse i mikroorganismer og planter, hindring af nitrogenoptagelse i sygdomsfremkaldende mikroorganismer, samt i forbindelse med sygdomme, der skyldes en nedsat transportfunktion af Rhesus-proteiner. Derfor er det vigtigt at forstå transportmekanismen i detalje, dvs. helt ned på atomart niveau at forstå sammenhængen mellem proteinstruktur og -funktion. I den forbindelse spiller proteinbevægelser (dynamik) en stor rolle, og MD simuleringer har mange gange tidligere vist sig som et værdifuldt værktøj til at opnå den ønskede forståelse.

Som den første AMT proteinstruktur blev røntgenkrystalstrukturen af AmtB betemt i 2004. Forståelse af sammenhængen mellem proteinstruktur og -funktion af et AMT protein kan forventes at give en generel forståelse af hele familien. Da MD simuleringer forudsætter god strukturel information, var det oplagt at undersøge disse sammenhænge vha. MD simuleringer af AmtB. Det blev derfor emnet for mit Ph.D.-projekt, som nu afsluttes med denne afhandling.

Nogle vigtige spørgsmål, som ønskes besvaret omkring funktionen af

AmtB, og AMT proteiner generelt, er bl.a. følgende: *1*) Hvordan rekrutterer AmtB substratet ammoniak/ammonium? *2*) Hvordan genkender AmtB substratet? Og da eksperimenter tyder på, at NH_4^+ rekrutteres, mens NH_3 transporteres, så *3*) hvor og hvordan deprotoneres NH_4^+?

Hovedresultaterne beskrevet i afhandlingen kan sammenfattes i en besvarelse på disse spørgsmål.

Hvordan rekrutterer AmtB substratet ammoniak/ammonium? Modsat tidligere fremsatte hypoteser om at aromatiske fenylalanin-sidekæder (F107 og W148) vha. såkaldte kation-π vekselvirkninger spiller en meget vigtig rolle i at rekruttere positivt ladet substrat, såsom NH_4^+, fremfor neutralt substrat, såsom NH_3, viste simuleringsresultaterne, at disse ikke har særlig stor betydning i så henseende. Modsat viste simuleringerne, at den deprotonerede asparaginsyre-sidekæde (Asp160), som tidligere kun var tillagt en proteinstruktur-stabiliserende funktion, derudover også har stor betydning for at rekruttere NH_4^+ fremfor NH_3. Simuleringerne afslørede samtidigt, at et karbonyloxygen (A162:O) er vigtig for rekrutteringen af NH_4^+ og videre transport.

Hvordan genkender AmtB substratet? Tidligere eksperimenter har indikeret at de fysiologisk vigtige natrium- og kaliumioner fravælges af AmtB og at deres tilstedeværelse ikke har betydning for rekrutteringen af NH_4^+. Det er interessant da kaliumionen er af samme størrelse og ladning som NH_4^+, og derfor skulle formodes at blive rekrutteret. Igen viste det sig A162:O spillede en vigtig rolle ved kun at binde kationer, som indeholder brint, men ikke metal ioner. Derudover viste simuleringerne, at en glutamin-sidekæde (Q104) kan stabilisere NH_4^+ ved at polarisere et vandmolekyle bundet til substratet, men muligvis også kan hindre større ioner såsom methylammonium ($CH_3NH_3^+$) i at binde ligeså godt som $NH4^+$.

Hvor og hvordan deprotoneres NH_4^+? Tidligere eksperimenter har vist, at den brintion (H^+), som spaltes fra når NH_4^+ deprotoneres, ikke transporteres igennem AmtB. På baggrund af indledende simuleringer blev der forslået en mekanisme, som kunne forklarer denne observation. Efterfølgende blev den estimerede position, hvor deprotonering finder sted, bekræftet vha. af energetiske beregninger baseret på såkaldte styrede MD simuleringer, hvor NH_3 og NH_4^+ blev trukket igennem kanalen. Selve mekanismen, som involverer ovenfor nævnte A162 og D160, blev dernæst testet vha. af en special type MD simulering, hvor dele af systemet behandles kvantemekanisk. Den foreslåede mekanisme kunne dog hverken be- eller afkræftes.

Bibliography

1. Solomon, E., L. Berg, and D. Martin. 1999. Biology. 5th edition. Saunders College Publishing, Orlando.

2. Kleiner, D. 1985. Bacterial ammonium transport. *FEMS Microbiology Letters* 32:87.

3. Kleiner, D. 1993. NH_4^+ Transport Systems. In Alkali Cation Transport Systems in Prokaryotes. E.P. Bakker, editor. CRC Press, Boca Raton. 379–396.

4. Ninnemann, O., J. C. Jauniaux, and W. B. Frommer. 1994. Identification of a high affinity NH_4^+ transporter from plants. *EMBO J.* 13:3464–3471.

5. Marini, A., S. Vissers, A. Urrestarazu, and B. Andre. 1994. Cloning and expression of the MEP1 gene encoding an ammonium transporter in *Saccharomyces cerevisiae*. *EMBO J.* 13:3456–3463.

6. Siewe, R. M., B. Weil, A. Burkovski, B. J. Eikmanns, M. Eikmanns, and R. Krämer. 1996. Functional and genetic characterization of the (methyl)ammonium uptake carrier of *Corynebacterium glutamicum*. *J. Biol. Chem.* 271:5398–5403.

7. Marini, A., A. Urrestarazu, R. Beauwens, and B. Andre. 1997. The Rh (Rhesus) blood group polypeptides are related to NH_4^+ transporters. *Trends in Biochemical Sciences* 22:460–461.

8. Transport Classification Database TCDB. 2005. `http://www.tcdb.org/`.

9. Murzin, A. G., S. E. Brenner, T. Hubbard, and C. Chothia. 1995. SCOP: A structural classification of proteins database for the investigation of sequences and structures. *J. Mol. Biol.* 247:536–540.

10. Saier Jr., M. H., B. H. Eng, S. Fard, J. Garg, D. A. Haggerty, W. J. Hutchinson, D. L. Jack, E. C. Lai, H. J. Liu, D. P. Nusinew, A. M. Omar, S. S. Pao, I. T. Paulsen, J. A. Quan, M. Sliwinski, T. Tseng,

S. Wachi, and G. B. Young. 1999. Phylogenetic characterization of novel transport protein families revealed by genome analyses. *Biochimica et Biophysica Acta (BBA) - Reviews on Biomembranes* 1422:1–56.

11. Merrick, M., A. Javelle, A. Durand, E. Severi, J. Thornton, N. D. Avent, M. J. Conroy, and P. A. Bullough. 2006. The *Escherichia coli* AmtB protein as a model system for understanding ammonium transport by Amt and Rh proteins. *Transfusion Clinique et Biologique* 13:97–102.

12. Khademi, S., and R. M. Stroud. 2006. The Amt/MEP/Rh family: Structure of AmtB and the mechanism of ammonia gas conduction. *Physiology* 21:419–429.

13. Ludewig, U., B. Neuhauser, and M. Dynowski. 2007. Molecular mechanisms of ammonium transport and accumulation in plants. *FEBS Letters* 581:2301–2308.

14. Winkler, F. 2006. Amt/MEP/Rh proteins conduct ammonia. *Pflügers Archiv European Journal of Physiology* 451:701–707.

15. Loque, D., and N. Von Wiren. 2004. Regulatory levels for the transport of ammonium in plant roots. *Journal of Experimental Botany* 55:1293–1305.

16. Huang, C., and J. Peng. 2005. Evolutionary conservation and diversification of Rh family genes and proteins. *PNAS* 102:15512–15517.

17. Khademi, S., J. O'Connell, J. Remis, Y. Robles-Colmenares, L. J. W. Miercke, and R. M. Stroud. 2004. Mechanism of ammonia transport by Amt/MEP/Rh: structure of AmtB at 1.35 å. *Science* 305:1587–1594.

18. Zheng, L., D. Kostrewa, S. Berneche, F. K. Winkler, and X. D. Li. 2004. The mechanism of ammonia transport based on the crystal structure of AmtB of *Escherichia coli*. *Proc. Natl. Acad. Sci. U. S. A.* 101:17090–17095.

19. Smart, O. S., J. M. Goodfellow, and B. A. Wallace. 1993. The pore dimensions of gramicidin. *Biophys. J.* 65:2455–2460.

20. Andrade, S. L. A., A. Dickmanns, R. Ficner, and O. Einsle. 2005. Crystal structure of the archaeal ammonium transporter Amt-1 from *Archaeoglobus fulgidus*. *Proc. Natl. Acad. Sci. U. S. A.* 102:14994–14999.

21. Conroy, M. J., A. Durand, D. Lupo, X. Li, P. A. Bullough, F. K. Winkler, and M. Merrick. 2007. The crystal structure of the *Escherichia coli* amtB-GlnK complex reveals how GlnK regulates the ammonia

channel. *Proceedings of the National Academy of Sciences of the USA* 104:1213–1218.

22. Gruswitz, F., J. O'Connell III, and R. M. Stroud. 2007. Inhibitory complex of the transmembrane ammonia channel, AmtB, and the cytosolic regulatory protein, GInK, at 1.96 å. *Proceedings of the National Academy of Sciences of the United States of America* 104:42–47.

23. Thomas, G., J. G. L. Mullins, and M. Merrick. 2000. Membrane topology of the Mep/Amt family of ammonium transporters. *Mol. Microbiol.* 37:331–344.

24. Monahan, B. J., S. E. Unkles, T. Tsing I, J. R. Kinghorn, M. J. Hynes, and M. A. Davis. 2002. Mutation and functional analysis of the *Aspergillus nidulans* ammonium permease MeaA and evidence for interaction with itself and MepA. *Fungal Genetics and Biology* 36:35–46.

25. Javelle, A., D. Lupo, L. Zheng, X. Li, F. K. Winkler, and M. Merrick. 2006. An Unusual Twin-His Arrangement in the Pore of Ammonia Channels Is Essential for Substrate Conductance. *J. Biol. Chem.* 281:39492–39498.

26. Meier-Wagner, J., L. Nolden, M. Jakoby, R. Siewe, R. Kramer, and A. Burkovski. 2001. Multiplicity of ammonium uptake systems in *Corynebacterium glutamicum*: role of Amt and AmtB. *Microbiology* 147:135–143.

27. Ludewig, U., N. von Wiren, and W. B. Frommer. 2002. Uniport of NH_4^+ by the root hair plasma membrane ammonium transporter LeAMT1;1. *J. Biol. Chem.* 277:13548–13555.

28. Ludewig, U. 2004. Electroneutral ammonium transport by basolateral Rhesus B glycoprotein. *J. Physiol. (Lond).* 559:751–759.

29. Soupene, E., L. He, D. Yan, and S. Kustu. 1998. Ammonia acquisition in enteric bacteria: Physiological role of the ammonium/methylammonium transport B (AmtB) protein. *Proc. Natl. Acad. Sci. U. S. A.* 95:7030–7034.

30. Soupene, E., R. M. Ramirez, and S. Kustu. 2001. Evidence that fungal MEP proteins mediate diffusion of the uncharged species NH_3 across the cytoplasmic membrane. *Mol. Cell. Biol.* 21:5733–5741.

31. Soupene, E., H. Lee, and S. Kustu. 2002. Ammonium/methylammonium transport (Amt) proteins facilitate diffusion of NH_3 bidirectionally. *Proc. Natl. Acad. Sci. U. S. A.* 99:3926–3931.

32. Soupene, E., T. Chu, R. W. Corbin, D. F. Hunt, and S. Kustu. 2002. Gas Channels for NH_3: Proteins from Hyperthermophiles Complement an *Escherichia coli* Mutant. *J. Bacteriol.* 184:3396–3400.

33. Ripoche, P., O. Bertrand, P. Gane, C. Birkenmeier, Y. Colin, and J.-P. Cartron. 2004. Human Rhesus-associated glycoprotein mediates facilitated transport of NH_3 into red blood cells. *Proc. Natl. Acad. Sci. U. S. A.* 101:17222–17227.

34. Javelle, A., G. Thomas, A.-M. Marini, R. Kramer, and M. Merrick. 2005. In vivo functional characterization of the *Escherichia coli* ammonium channel AmtB: Evidence for metabolic coupling of AmtB to glutamine synthetase. *Biochem. J.* 390:215–222.

35. Javelle, A., D. Lupo, X. D. Li, M. Merrick, M. Chami, P. Ripoche, and F. K. Winkler. 2007. Structural and mechanistic aspects of Amt/Rh proteins. *Journal of Structural Biology* 158:472–481.

36. Mayer, M., G. Schaaf, I. Mouro, C. Lopez, Y. Colin, P. Neumann, J. Cartron, and U. Ludewig. 2006. Different Transport Mechanisms in Plant and Human AMT/Rh-type Ammonium Transporters. *J. Gen. Physiol.* 127:133–144.

37. Mayer, M., M. Dynowski, and U. Ludewig. 2006. Ammonium ion transport by the AMT/Rh homologue LeAMT1;1. *Biochemical Journal* 396:431–437.

38. Wood, C. C., F. Poree, I. Dreyer, G. J. Koehler, and M. K. Udvardi. 2000. Mechanisms of ammonium transport, accumulation, and retention in ooyctes and yeast cells expressing arabidopsis atamt1;1. *FEBS Letters* 580:3931–3936.

39. Hackette, S. L., G. E. Skye, C. Burton, and I. H. Segel. 1970. Characterization of an ammonium transport system in filamentous fungi with methylammonium-14C as the substrate. *J. Biol. Chem.* 245:4241–4250.

40. Marini, A.-M., G. Matassi, V. Raynal, B. Andre, J.-P. Cartron, and B. Cherif-Zahar. 2000. The human Rhesus-associated RhAG protein and a kidney homologue promote ammonium transport in yeast. *Nat. Genet.* 26:341–344.

41. Soupene, E., N. King, E. Feild, P. Liu, K. K. Niyogi, C.-H. Huang, and S. Kustu. 2002. Rhesus expression in a green alga is regulated by CO_2. *Proc. Natl. Acad. Sci. U. S. A.* 99:7769–7773.

42. Soupene, E., W. Inwood, and S. Kustu. 2004. From The Cover: Lack of the Rhesus protein Rh1 impairs growth of the green alga *Chlamydomonas reinhardtii* at high CO_2. *PNAS* 101:7787–7792.

43. Schneider, M., I. W. Marison, and U. von Stockar. 1996. The importance of ammonia in mammalian cell culture. *J. Biotechnol.* 46:161–185.

44. Knepper, M. A., R. Packer, and D. W. Good. 1989. Ammonium transport in the kidney. *Physiol. Rev.* 69:179–249.

45. Olde Damink, S. W. M., N. E. P. Deutz, C. H. C. Dejong, P. B. Soeters, and R. Jalan. 2002. Interorgan ammonia metabolism in liver failure. *Neurochemistry International* 41:177–188.

46. Haussinger, D., W. H. Lamers, and A. F. M. Moorman. 1992. Hepatocyte heterogeneity in the metabolism of amino-acids and ammonia. *Enzyme* 46:72–93.

47. Ott, P., and F. S. Larsen. 2004. Blood-brain barrier permeability to ammonia in liver failure: a critical reappraisal. *Neurochem. Int.* 44:185–198.

48. Felipo, V., and R. F. Butterworth. 2002. Neurobiology of ammonia. *Prog. Neurobiol.* 67:259–279.

49. Shawcross, D. L., S. W. M. Olde Damink, R. F. Butterworth, and R. Jalan. 2005. Ammonia and hepatic encephalopathy: The more things change, the more they remain the same. *Metabolic Brain Disease* 20:169–179.

50. Labow, B. I., W. W. Souba, and S. F. Abcouwer. 1999. Glutamine synthetase expression in muscle is regulated by transcriptional and posttranscriptional mechanisms. *Am J Physiol Endocrinol Metab* 276:E1136–1145.

51. Nakhoul, N. L., and L. L. Hamm. 2004. Non-erythroid Rh glycoproteins: a putative new family of mammalian ammonium transporters. *Pflügers Archiv European Journal of Physiology* 447:807–812.

52. Conroy, M. J., P. A. Bullough, M. Merrick, and N. D. Avent. 2005. Modelling the human rhesus proteins: implications for structure and function. *British Journal of Haematology* 131:543–551.

53. Zidi-Yahiaoui, N., I. Mouro-Chanteloup, A. M. D'Ambrosio, C. Lopez, P. Gane, C. L. V. Kim, J. P. Cartron, Y. Colin, and P. Ripoche. 40. Human Rhesus B and Rhesus C glycoproteins: properties of facilitated ammonium transport in recombinant kidney cells. *Biochemical Journal* 391:33.

54. Planelles, G. 2007. Ammonium homeostasis and human rhesus glycoproteins. *Nephron Physiology* 105:11–17.

55. Helling, R. B. 1998. Pathway Choice in Glutamate Synthesis in *Escherichia coli*. *J. Bacteriol.* 180:4571–4575.

56. Kabir, M. S., T. Sagara, T. Oshima, Y. Kawagoe, H. Mori, R. Tsunedomi, and M. Yamada. 2004. Effects of mutations in the rpos gene on cell viability and global gene expression under nitrogen starvation in *Escherichia coli*. *Microbiology (Read)* 150:2543–2554.

57. vanHeeswijk, W. C., S. Hoving, D. Molenaar, B. Stegeman, D. Kahn, and H. V. Westerhoff. 1996. An alternative P-ii protein in the regulation of glutamine synthetase in *Escherichia coli*. *Mol. Microbiol.* 21:133–146.

58. Thomas, G., G. Coutts, and M. Merrick. 2000. The glnKamtB operon: a conserved gene pair in prokaryotes. *Trends in Genetics* 16:11–14.

59. Javelle, A., E. Severi, J. Thornton, and M. Merrick. 2004. Ammonium sensing in *Escherichia coli*: Role of the ammonium transporter Amtb and AmtB-GlnK complex formation. *J. Biol. Chem.* 279:8530–8538.

60. Durand, A., and M. Merrick. 2006. In Vitro Analysis of the *Escherichia coli* AmtB-GlnK Complex Reveals a Stoichiometric Interaction and Sensitivity to ATP and 2-Oxoglutarate. *J. Biol. Chem.* 281:29558–29567.

61. Barnes, Jr.and A. Jayakumar, E. M. 1993. NH_4^+ Transport Systems in *Escherichia coli*. In Alkali Cation Transport Systems in Prokaryotes. E.P. Bakker, editor. CRC Press, Boca Raton. 379–396.

62. Strosser, J., A. Schaffer, S. Kramer, and A. Burkovski. 2004. Regulation of GlnK activity: modification, membrane sequestration and proteolysis as regulatory principles in the network of nitrogen control in *Corynebacterium glutamicum*. *Mol. Microbiol.* 54:132–147.

63. Coutts, G., T. Thomas, D. Blakey, and M. Merrick. 2002. Membrane sequestration of the signal transduction protein GlnK by the ammonium transporter AmtB. *EMBO J.* 21:536–545.

64. Yakunin, A. F., and P. C. Hallenbeck. 2002. AmtB is necessary for NH_4^+-induced nitrogenase switch-off and ADP-ribosylation in *Rhodobacter capsulatus*. *J. Bacteriol.* 184:4081–4088.

65. Javelle, A., B. Andre, A. Marini, and M. Chalot. 2003. High-affinity ammonium transporters and nitrogen sensing in *mycorrhizas*. *Trends in Microbiology* 11:53–55.

66. Dixon, R., and D. Kahn. 2004. Genetic regulation of biological nitrogen fixation. *Nat. Rev. Microbiol.* 2:621–631.

67. Javelle, A., and M. Merrick. 2005. Complex formation between AmtB and GlnK: an ancestral role in prokaryotic nitrogen control. *Biochem. Soc. Trans* 33:170–172.

68. Smith, D. G., M. D. Garcia-Pedrajas, S. E. Gold, and M. H. Perlin. 2003. Isolation and characterization from pathogenic fungi of genes encoding ammonium permeases and their roles in dimorphism. *Molecular Microbiology* 50:259–275.

69. Madhani, H. D., and G. R. Fink. 1998. The control of filamentous differentiation and virulence in fungi. *Trends in Cell Biology* 8:348–353.

70. Conroy, M. J., S. J. Jamieson, D. Blakey, T. Kaufmann, A. Engel, D. Fotiadis, M. Merrick, and P. A. Bullough. 2004. Electron and atomic force microscopy of the trimeric ammonium transporter AmtB. *EMBO Rep.* 5:1153–1158.

71. Blakey, D., A. Leech, G. H. Thomas, G. Coutts, K. Findlay, and M. Merrick. 2002. Purification of the *Escherichia coli* ammonium transporter AmtB reveals a trimeric stoichiometry. *Biochem. J.* 364:527–536.

72. Kumar, A., B. N. Kaiser, M. Y. Siddiqi, and A. D. M. Glass. 2006. Functional characterisation of OsAMT1.1 overexpression lines of rice, *Oryza sativa*. *Functional Plant Biology* 33:339–346.

73. Liu, Y., and X. Hu. 2006. Molecular determinants for binding of ammonium ion in the ammonia transporter AmtB - a quantum chemical analysis. *J. Phys. Chem. A* 110:1375–1381.

74. Luzhkov, V. B., M. Almlof, M. Nervall, and J. Åqvist. 2006. Computational study of the binding affinity and selectivity of the bacterial ammonium transporter AmtB. *Biochemistry* 45:10807–10814.

75. Nygaard, T. P., C. Rovira, G. H. Peters, and M. O. Jensen. 2006. Ammonium recruitment and ammonia transport by *E. coli* ammonia channel AmtB. *Biophys. J.* 91:4401–4412.

76. Lin, Y., Z. Cao, and Y. Mo. 2006. Molecular dynamics simulations on the *Escherichia coli* ammonia channel protein AmtB: Mechanism of ammonia/ ammonium transport. *Journal of the American Chemical Society* 128:10876–10884.

77. Ishikita, H., and E. Knapp. 2007. Protonation states of ammonia/ammonium in the hydrophobic pore of ammonia transporter protein AmtB. *Journal of the American Chemical Society* 129:1210–1215.

78. Yang, H., Y. Xu, W. Zhu, K. Chen, and H. Jiang. 2007. Detailed Mechanism for AmtB Conducting NH_4^+/NH_3: Molecular Dynamics Simulations. *Biophys. J.* 92:877–885.

79. Lamoureux, G., M. L. Klein, and S. Berneche. 2007. A stable water chain in the hydrophobic pore of the AmtB ammonium transporter. *Biophys. J.* 92:L82–84.

80. Bostick, D. L., and C. L. Brooks III. 2007. Deprotonation by dehydration: The origin of ammonium sensing in the AmtB channel. *PLoS Computational Biology* 3:0231–0246.

81. Allen, M. P., and D. J. Tildesley. 1987. Computer simulation of liquids. 6th edition. Oxford Clarendon Pres.

82. Frenkel, D., and B. Smit. 1996. Understanding Molecular Simulation - From Algorithms to Applications. 1st edition. Academic Press, San Diego, California.

83. Leach, A. R. 2001. Molecular Modelling: principles and applications. 2nd edition. Prentice Hall.

84. Rapaport, D. C. 1995. The art of molecular dynamics simulation. Cambridge University Press.

85. Haile, J. M. 1997. Molecular dynamics simulation: elementary methods. Wiley.

86. Swope, W. C., H. C. Andersen, P. H. Berens, and K. R. Wilson. 1982. A computer simulation method for the calculation of equilibrium constants for the formation of physical clusters of molecules: Application to small water clusters. *The Journal of Chemical Physics* 76:637–649.

87. Bhandarkar, M., R. Brunner, C. Chipot, A. Dalke, S. Dixit, P. Grayson, J. Gullingsrud, A. Gursoy, D. Hardy, J. Hénin, W. Humphrey, D. Hurwitz, N. Krawetz, S. Kumar, M. Nelson, J. Phillips, A. Shinozaki, G. Zheng, and F. Zhu. 2006. NAMD User's Guide - Version 2.6. Theoretical Biophysics Group, Beckman Institute, University of Illinois, IL, Urbana.

88. Marx, D., and J. Hutter. 2000. Ab initio molecular dynamics: Theory and Implementation in Modern Methods and Algorithms of Quantum Chemistry. John von Neumann Institute for Computing, Julich.

89. Tuckerman, M., B. J. Berne, and G. J. Martyna. 19920801. Reversible multiple time scale molecular dynamics. *Journal of Chemical Physics* 97:1990–2001.

90. Verlet, L. 1967. Computer "experiments" on classical fluids. i. thermodynamical properties of lennard-jones molecules. *Phys. Rev.* 159:98–103.

91. Hockney, R. W., and J. W. Eastwood. 1981. Computer simulation using particles. McGraw-Hill.

92. Auerbach, D. J., W. Paul, A. F. Bakker, C. Lutz, W. E. Rudge, and F. F. Abraham. 1987. A special purpose parallel computer for molecular dynamics: motivation, design, implementation, and application. *Journal of Physical Chemistry* 91:4881–4890.

93. Ewald, P. P. 1921. Die berechnung optischer und elektrostatischer gitterpotentiale. *Annalen der Physik* 369:253–287.

94. Darden, T., D. York, and L. Pedersen. 1993. Particle mesh ewald: An nÂ·log(n) method for ewald sums in large systems. *Journal of Chemical Physics* 98:10089–10092.

95. Eastwood, J. W., and R. W. Hockney. 1974. Shaping the force law in two-dimensional particle-mesh models. *Journal of Computational Physics* 16:342–359.

96. Feller, S. E., Y. H. Zhang, R. W. Pastor, and B. R. Brooks. 1995. Constant-pressure molecular-dynamics simulation - the langevin piston method. *Journal of Chemical Physics* 103:4613–4621.

97. Ponder, J. W., and D. A. Case. 2003. Force fields for protein simulations. *Advances in Protein Chemistry* 66:27–85.

98. Patel, S., and C. L. Brooks. 2006. Fluctuating charge force fields: Recent developments and applications from small molecules to macromolecular biological systems. *Molecular Simulation* 32:231–249.

99. Duan, Y., C. Wu, S. Chowdhury, M. C. Lee, G. Xiong, W. Zhang, R. Yang, P. Cieplak, R. Luo, T. Lee, J. Caldwell, J. Wang, and P. Kollman. 2003. A point-charge force field for molecular mechanics simulations of proteins based on condensed-phase quantum mechanical calculations. *Journal of Computational Chemistry* 24:1999–2012.

100. van Duin, A. C. T., S. Dasgupta, F. Lorant, and W. A. Goddard. 2001. ReaxFF: A reactive force field for hydrocarbons. *Journal of Physical Chemistry A* 105:9396–9409.

101. Car, R., and M. Parrinello. 1985. Unified approach for molecular dynamics and density-functional theory. *Phys. Rev. Lett.* 55:2471–2474.

102. Remler, D. K., and P. A. Madden. 1990. Molecular-dynamics without effective potentials via the Car-Parrinello approach. *Molecular Physics* 70:921–966.

103. Lin, H., and D. G. Truhlar. 2007. QM/MM: what have we learned, where are we, and where do we go from here? *Theoretical Chemistry Accounts* 117:185–199.

104. van Gunsteren, W. F., S. R. Billeter, A. A. Eising, P. H. Hünenberger, P. Krüger, A. E. Mark, W. R. P. Scott, and I. G. Tironi. 1996. Biomolecular Simulation: The GROMOS96 Manual and User Guide. Vdf Hochschulverlag AG an der ETH Zürich. Zürich.

105. Phillips, J. C., R. Braun, W. Wang, J. Gumbart, E. Tajkhorshid, E. Villa, C. Chipot, R. D. Skeel, L. Kale, and K. Schulten. 2005. Scalable molecular dynamics with NAMD. *J. Comput. Chem.* 26:1781–1802, `http://www.ks.uiuc.edu/Research/namd/`.

106. Case, D. A., T. E. Cheatham III, T. Darden, H. Gohlke, R. Luo, K. M. Merz Jr., A. Onufriev, C. Simmerling, B. Wang, and R. J. Woods. 2005. The Amber biomolecular simulation programs. *Journal of Computational Chemistry* 26:1668–1688.

107. Brooks, B. R., R. E. Bruccoleri, B. D. Olafson, D. J. States, S. Swaminathan, and M. Karplus. 1983. CHARMM: a program for macromolecular energy, minimisation, and dynamics calculations. *J. Comput. Chem.* 4:187–217.

108. Van Der Spoel, D., E. Lindahl, B. Hess, G. Groenhof, A. E. Mark, and H. J. C. Berendsen. 2005. GROMACS: Fast, flexible, and free. *Journal of Computational Chemistry* 26:1701–1718.

109. Eichinger, M., P. Tavran, J. Hutter, and M. Parrinello. 1999. A hybrid method for solutes in complex solvents: Density functional theory combined with empirical force fields. *J. Chem. Phys.* 110:10452–10467.

110. Laio, A., J. VandeVondele, and U. Röthlisberger. 2002. A hamiltonian electrostatic coupling scheme for hybrid Car-Parrinello molecular dynamics simulations. *J. Chem. Phys.* 116:6941–6947.

111. Copyright IBM Corp 1990-2006, C. M. f. F. S. .-. 1990–2006. CPMD v3.11. http://www.cpmd.org.

112. Sanbonmatsu, K., and C. Tung. 2007. High performance computing in biology: Multimillion atom simulations of nanoscale systems. *Journal of Structural Biology* 157:470–480.

113. Seibert, M. M., A. Patriksson, B. Hess, and D. van der Spoel. 2005. Reproducible polypeptide folding and structure prediction using molecular dynamics simulations. *Journal of Molecular Biology* 354:173–183.

114. Monticelli, L., D. P. Tieleman, and G. Colombo. 2005. Mechanism of helix nucleation and propagation: Microscopic view from microsecond time scale md simulations. *Journal of Physical Chemistry B* 109:20064–20067.

115. Park, S., F. Khalili-Araghi, E. Tajkhorshid, and K. Schulten. 2003. Free energy calculation from steered molecular dynamics simulations using Jarzynski's equality. *J. Chem. Phys.* 119:3559–66.

116. Jensen, M. O., S. Park, E. Tajkhorshid, and K. Schulten. 2002. Energetics of glycerol conduction through aquaglyceroporin GlpF. *Proc. Natl. Acad. Sci. U. S. A.* 99:6731–6736.

117. Hwang, H., G. C. Schatz, and M. A. Ratner. 2006. Steered molecular dynamics studies of the potential of mean force of a Na^+ or K^+ ion in a cyclic peptide nanotube. *Journal of Physical Chemistry B* 110:26448–26460.

118. Liu, Z., Y. Xu, and P. Tang. 2006. Steered molecular dynamics simulations of Na^+ permeation across the gramicidin A channel. *Journal of Physical Chemistry B* 110:12789–12795.

119. Rovira, C. 2003. The structure and dynamics of the Fe-Co bond in myoglobin. *Journal of Physics: Condensed Matter* 15:S1809–S1822.

120. Spiegel, K., and A. Magistrato. 2006. Modeling anticancer drug-DNA interactions via mixed QM/MM molecular dynamics simulations. *Organic and Biomolecular Chemistry* 4:2507–2517.

121. Mulholland, A. J. 2005. Modelling enzyme reaction mechanisms, specificity and catalysis. *Drug Discovery Today* 10:1393–1402.

122. Colombo, M. C., L. Guidoni, A. Laio, A. Magistrato, P. Maurer, S. Piana, U. Rohrig, K. Spiegel, M. Sulpizi, J. VandeVondele, M. Zumstein, and U. Röthlisberger. 2002. Hybrid QM/MM Car-Parrinello simulations of catalytic and enzymatic reactions. *Chimia* 56:13–19.

123. Senn, H. M., and W. Thiel. 2007. QM/MM methods for biological systems. *Topics in Current Chemistry* 268:173–290.

124. McQuarrie, D. 2000. Statistical Mechanics. 1st edition. University Science Books.

125. van Gunsteren, W. F., T. C. Beutler, F. Fraternali, P. M. King, A. E. Mark, and P. E. Smith. 1993. Computation of free energy in practice: Choice of approximation and accuracy limiting factors. *In* Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. Volume 2., W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors. ESCOM Science Publishers B.V., Leiden, The Netherlands, 315–347.

126. King, P. M. 1993. Free energy via molecular simulation: A primer. *In* Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. Volume 2., W. F. van Gunsteren, P. K. Weiner, and A. J. Wilkinson, editors. ESCOM Science Publishers B.V., Leiden, The Netherlands, 267–314.

127. Beveridge, D. L., and F. M. DiCapua. 1989. Free-energy via molecular simulation - applications to chemical and biomolecular systems. *Annual Review of Biophysics and Biophysical Chemistry* 18:431–492.

128. Park, S., and K. Schulten. 2004. Calculating potentials of mean force from steered molecular dynamics simulations. *J. Chem. Phys.* 120:5946–5961.

129. van Gunsteren, W. F., and P. K. Weiner, editors. 1989. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. ESCOM Science Publishers B.V., Leiden, The Netherlands.

130. van Gunsteren, W. F., P. K. Weiner, and A. J. Wilkinson, editors. 1993. Computer Simulation of Biomolecular Systems: Theoretical and Experimental Applications. Volume 2. ESCOM Science Publishers B.V., Leiden, The Netherlands.

131. Straatsma, T. P., and J. A. McCammon. 1992. Computational alchemy. *Annual Review of Physical Chemistry* 43:407–435.

132. Zwanzig, R. W. 1954. High-temperature equation of state by a perturbation method. I. Nonpolar gases. *The Journal of Chemical Physics* 22:1420–1426.

133. Kirkwood, J. G. 1935. Statistical mechanics of fluid mixtures. *The Journal of Chemical Physics* 3:300–313.

134. Torrie, G. M., and J. P. Valleau. 1977. Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling. *Journal of Computational Physics* 23:187–199.

135. Crooks, G. 2000. Path-ensemble averages in systems driven far from equilibrium. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 61:2361–6.

136. Crooks, G. 1999. Entropy production fluctuation theorem and the nonequilibrium work relation for free energy differences. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 60:2721–6.

137. Tolman, R. 1979. The Principles of Statistical Mechanics. Dover Publications, Inc., New York, New York.

138. van Kampen, N. 1983. Stochastic Processes in Physics and Chemistry. 1st edition. NorthH̃olland.

139. Jarzynski, C. 1997. Equilibrium free-energy differences from nonequilibrium measurements: a master-equation approach. *Phys. Rev. E Stat. Nonlin. Soft Matter Phys.* 56:5018–35.

140. Jarzynski, C. 1997. Nonequilibrium equality for free energy differences. *Phys. Rev. Lett.* 78:2690–3.

141. Hummer, G. 2001. Fast-growth thermodynamic integration: Error and efficiency analysis. *J. Chem. Phys.* 114:7330–7337.

142. Chandler, D. 1987. Introduction to Modern Statistical Mechanics. 1st edition. Oxford University Press, Inc., New York, New York.

143. Trzesniak, D., A. E. Kunz, and W. F. van Gunsteren. 2007. A comparison of methods to compute the potential of mean force. *ChemPhysChem* 8:162–169.

144. Kumar, S., J. M. Rosenberg, D. Bouzida, R. H. Swendsen, and P. A. Kollman. 1992. The weighted histogram analysis method for free-energy calculations on biomolecules. I. The method. *Journal of Computational Chemistry* 13:1011–1021.

145. Kosztin, I., B. Barz, and L. Janosi. 2006. Calculating potentials of mean force and diffusion coefficients from nonequilibrium processes without Jarzynski's equality. *J. Chem. Phys.* 124:064106–1–064106–11.

146. Hummer, G., and A. Szabo. 2001. Free energy reconstruction from nonequilibrium single-molecule pulling experiments. *Proc. Natl. Acad. Sci. U. S. A.* 98:3658–3661.

147. Hummer, G., and A. Szabo. 2005. Free energy surfaces from single-molecule force spectroscopy. *Accounts of Chemical Research* 38:504–513.

148. Weisstein, E. W. 2007. "Boxcar Function." From MathWorld – A Wolfram Web Resource.
Website: `http://mathworld.wolfram.com/BoxcarFunction.html`.

149. Balsera, M., S. Stepaniants, I. S, Y. Oono, and K. Schulten. 1997. Reconstructing potential energy functions from simulated force-induced unbinding processes. *Biophys. J.* 73:1281–1287.

150. Fisz, M. 1963. Probability Theory and Mathematical Statistics. John Wiley & Sons, New York.

151. Kale, L., K. Schulten, R. Skeel, M. Bhandarkar, R. Brunner, A. Gursoy, N. Krawetz, J. Phillips, A. Shinozaki, and K. Varadarajan. 1999. NAMD2: Greater scalability for parallel molecular dynamics. *J. Comput. Phys.* 151:283–312.

152. Humphrey, W., A. Dalke, and K. Schulten. 1996. VMD: Visual molecular dynamics. *J. Mol. Graph.* 14:33–38.

153. Tieleman, D. P., and H. J. C. Berendsen. 1998. A molecular dynamics study of the pores formed by *Escherichia coli* OmpF porin in a fully hydrated palmitoyloleoylphosphatidylcholine bilayer. *Biophys. J.* 74:2786–801.

154. Rappolt, M., A. Hickel, R. Bringezu, and K. Lohner. 2003. Mechanism of the lamellar/inverse hexagonal phase transition examined by high resolution x-ray diffraction. *Biophys. J.* 84:3111–3122.

155. SOLVATE v1.0. 1996. H. Grubmüller, Theoretical Biophysics Group, Institut für Medizinische Optik, Ludwig-Maximilians-Universität München, München, Germany.

156. MacKerell Jr., A. D., D. Bashford, M. Bellott, R. L. Dunbrack Jr., J. D. Evanseck, M. J. Field, S. Fischer, J. Gao, H. Guo, S. Ha, D. Joseph-McCarthy, L. Kuchnir, K. Kuczera, F. T. K. Lau, C. Mattos, S. Michnick, T. Ngo, D. T. Nguyen, B. Prodhom, W. E. R. III, B. Roux, M. Schlenkrich, J. C. Smith, R. Stote, J. Straub, M. Watanabe, J. Wiorkiewicz-Kuczera, D. Yin, and M. Karplus. 1998. All-atom empirical potential for molecular modeling and dynamics studies of proteins. *J. Phys. Chem. B* 102:3586–3616.

157. Chen, I. J., D. Yin, and A. D. Mackerell Jr. 2002. Combined ab initio/empirical approach for optimization of Lennard-Jones parameters for polar-neutral compounds. *J. Comput. Chem.* 23:199–213.

158. Jorgensen, W. L., and J. Gao. 1986. Monte Carlo simulations of the hydration of ammonium and carboxylate ions. *J. Phys. Chem.* 90:2174–2182.

159. Faraldo-Gomez, J. D., and B. Roux. 2004. Electrostatics of ion stabilization in a ClC chloride channel homologue from *Escherichia coli*. *J. Mol. Biol.* 339:981–1000.

160. Caldwell, J. W., and P. A. Kollman. 1995. Cation-π interactions: Nonadditive effects are critical in their accurate representation. *Journal of the American Chemical Society* 117:4177–4178.

161. Kaback, H. R. 2005. Structure and mechanism of the lactose permease. *C. R. Biol.* 328:557–567.

162. Karim, O. A., and A. D. J. Haymet. 1990. Dynamics of an ammonium ion in water - molecular-dynamics simulation. *J. Chem. Phys.* 93:5961–5966.

163. Agmon, N. 1995. The Grotthuss mechanism. *Chem. Phys. Lett.* 224:456–462.

164. Gouaux, E., and R. MacKinnon. 2005. Principles of selective ion transport in channels and pumps. *Science* 310:1461–1465.

165. Perrin, C. L. 1989. Proton exchange in amides: Surprises from simple systems. *Acc. Chem. Res.* 22:268–275.

166. Dempsey, C. E. 2001. Hydrogen exchange in peptides and proteins using NMR spectroscopy. *Prog. Nucl. Magn. Reson. Spectrosc.* 39:135–170.

167. Yoshikawa, S., K. Shinzawa-Itoh, R. Nakashima, R. Yaono, E. Yamashita, N. Inoue, M. Yao, M. J. Fei, C. P. Libeu, T. Mizushima, H. Yamaguchi, T. Tomizaki, and T. Tsukihara. 1998. Redox-Coupled Crystal Structural Changes in Bovine Heart Cytochrome c Oxidase. *Science* 280:1723–1729.

168. Tsukihara, T., K. Shimokata, Y. Katayama, H. Shimada, K. Muramoto, H. Aoyama, M. Mochizuki, K. Shinzawa-Itoh, E. Yamashita, M. Yao, Y. Ishimura, and S. Yoshikawa. 2003. The low-spin heme of cytochrome c oxidase as the driving element of the proton-pumping process. *PNAS* 100:15304–15309.

169. Isralewitz, B., M. Gao, and K. Schulten. 2001. Steered molecular dynamics and mechanical functions of proteins. *Curr. Opin. Struct. Biol.* 11:224–230.

170. Bashford, D., and M. Karplus. 1990. pK_a's of ionizable groups in proteins: atomic detail from a continuum electrostatic model. *Biochemistry* 29:10219–10225.

171. Labotka, R. J., P. Lundberg, and P. W. Kuchel. 1995. Ammonia permeability of erythrocyte membrane studied by 14N and 15N saturation transfer NMR spectroscopy. *Am. J. Physiol. Cell. Physiol.* 268:C686–699.

172. Montanini, B., N. Moretto, E. Soragni, R. Percudani, and S. Ottonello. 2002. A high-affinity ammonium transporter from the *mycorrhizal ascomycete Tuber borchii*. *Fungal Genetics and Biology* 36:22–34.

173. Marini, A., S. Soussi-Boudekou, S. Vissers, and B. Andre. 1997. A family of ammonium transporters in *Saccharomyces cerevisiae*. *Mol. Cell. Biol.* 17:4282–4293.

174. Jensen, M. O., U. Röthlisberger, and C. Rovira. 2005. Hydroxide and proton migration in aquaporins. *Biophys. J.* 89:1744–1759.

175. Cornell, W. E., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1995. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 117:5179–5197.

176. Cornell, W. D., P. Cieplak, C. I. Bayly, I. R. Gould, K. M. Merz Jr., D. M. Ferguson, D. C. Spellmeyer, T. Fox, J. W. Caldwell, and P. A. Kollman. 1996. A second generation force field for the simulation of proteins, nucleic acids, and organic molecules. *J. Am. Chem. Soc.* 118:2309.

177. Case, D. A., D. A. Pearlman, J. W. Caldwell, T. E. Cheatham III, J. Wang, W. S. Ross, C. L. Simmerling, T. A. Darden, K. M. Merz, R. V. Stanton, A. L. Cheng, J. J. Vincent, M. Crowley, V. Tsui, H. Gohlke, R. J. Radmer, Y. Duan, J. Pitera, I. Massova, G. L. Seibel, U. C. Singh, P. K. Weiner, and P. A. Kollman. 2002. AMBER 7, University of California, San Francisco.

178. Zhang, Y., T. Lee, and W. Yang. 1999. A pseudobond approach to combining quantum mechanical and molecular mechanical methods. *Journal of Chemical Physics* 110:46–54.

179. Carloni, P., U. Röthlisberger, and M. Parrinello. 2002. The role and perspective of *Ab Initio* molecular dynamics in the study of biological systems. *Accounts of Chemical Research* 35:455–464.

180. Blumberger, J., and M. L. Klein. 2006. Reorganization free energies for long-range electron transfer in a porphyrin-binding four-helix bundle protein. *Journal of the American Chemical Society* 128:13854–13867.

181. Biarnes, X., J. Nieto, A. Planas, and C. Rovira. 2006. Substrate distortion in the Michaelis complex of bacillus 1,3-1,4-beta-glucanase: insight from first principles molecular dynamics simulations. *J. Biol. Chem.* 281:1432–1441.

182. Dal Peraro, M., P. Ruggerone, S. Raugei, F. L. Gervasio, and P. Carloni. 2007. Investigating biological systems using first principles Car-Parrinello molecular dynamics simulations. *Current Opinion in Structural Biology* 17:149–156.

183. Bucher, D., S. Raugei, L. Guidoni, M. Dal Peraro, U. Röthlisberger, P. Carloni, and M. L. Klein. 2006. Polarization effects and charge transfer in the KcsA potassium channel. *Biophysical Chemistry* 124:292–301.

184. Kohn, W., and L. J. Sham. 1965. Self-consistent equations including exchange and correlation effects. *Phys. Rev.* 140:A1133–A1138.

185. Perdew, J. P., K. Burke, and M. Ernzerhof. 1996. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 77:3865–3868.

186. Perdew, J. P., K. Burke, and M. Ernzerhof. 1997. Generalized gradient approximation made simple. *Phys. Rev. Lett.* 78:1396.

187. Troullier, N., and J. L. Martins. 1991. Efficient pseudopotentials for plane-wave calculations. *Phys. Rev. B* 43:1993–2006.

188. Kleinman, L., and D. M. Bylander. 1982. Efficious form for model pseudopotentials. *Phys. Rev. Lett.* 48:1425–1428.

189. Martyna, G. J., and M. E. Tuckerman. 1999. A reciprocal space based method for treating long range interactions in *ab initio* and force-field-based calculations in clusters. *Journal of Chemical Physics* 110:2810–2821.

190. von Lilienfeld, O. A., I. Tavernelli, U. Röthlisberger, and D. Sebastiani. 2005. Variational optimization of effective atom centered potentials for molecular properties. *Journal of Chemical Physics* 122:014113/1–014113/6.

191. Ikeda, T., M. Boero, and K. Terakura. 2007. Hydration of alkali ions from first principles molecular dynamics revisited. *Journal of Chemical Physics* 126:34501.

192. Brugé, F., M. Bernasconi, and M. Parrinello. 1999. *Ab Initio* simulation of rotational dynamics of solvated ammonium ion in water. *Journal of the American Chemical Society* 121:10883–10888.

193. Reddy, A. S., and G. N. Sastry. 2005. Cation [M = H^+/, Li^+/, Na^+/, K^+/, Ca^{2+}/, Mg^{2+}/, NH_4^+/, and NMe_4^+/] interactions with the aromatic motifs of naturally occurring amino acids: A theoretical study. *Journal of Physical Chemistry A* 110:8893–8903.

194. Ben-Naim, A., and Y. Marcus. 1984. Solvation thermodynamics of nonionic solutes. *The Journal of Chemical Physics* 81:2016–2027.

195. Klots, E. C. 1981. Solubility of protons in water. *Journal of Physical Chemistry* 85:3585–3588.

196. Payzant, J., A. Cunningham, and P. Kebarle. 1973. Gas phase solvation of the ammonium ion by NH_3 and H_2O and stabilities of mixed clusters $NH_4^+(NH_3)_n(H_2O)_w$. *Can. J. Chem.* 51:3242–3249.

197. Dzidic, I., and P. Kebarle. 1970. Hydration of the alkali ions in the gas phase. enthalpies and entropies of reactions $M^+(H_2O)_{n-1}+H_2O=M^+(H_2O)_n$. *Journal of Physical Chemistry* 74:1466–1474.

Appendix A

Supplementary Figures to Chapter 6

A1

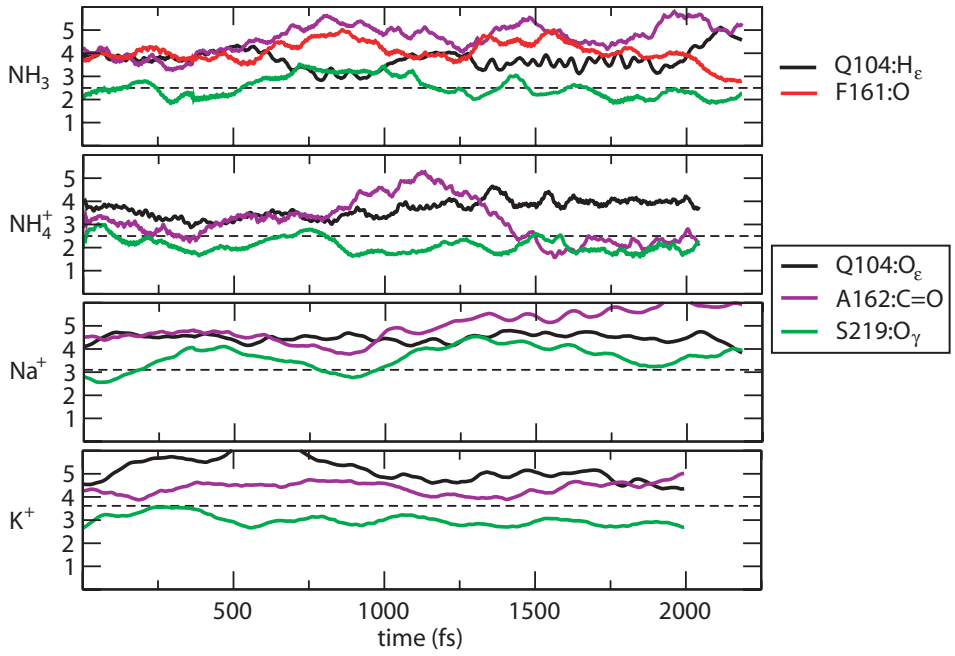

Figure A1: Distances between substrates and nearby residues at the Am1 position. For each substrate, the shortest distances to residues A162, S219, and Q104 are shown. For NH_3/NH_4^+ the shortest H-bond distances are reported. For NH_3 the shortest distance to Q104 is to H_ϵ, while for the other substrates the shortest distance is to O_ϵ. Since F161:O comes within 3 Å of NH_3 during simulation, that distance is also shown. The dashed line is the cutoff distance used for coordination number calculations.

A2

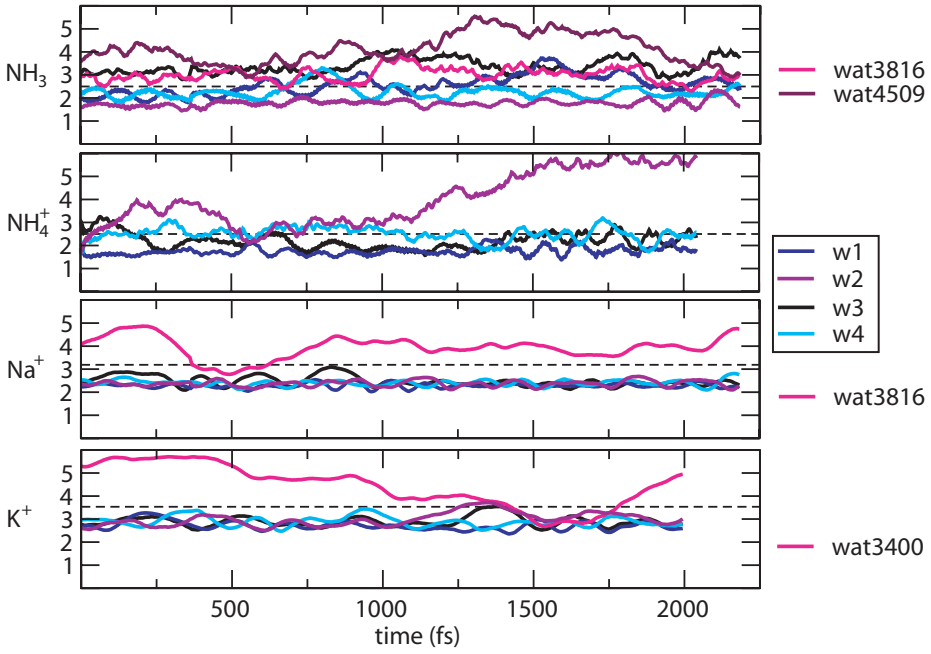

Figure A2: Distance between substrates and nearby water molecules at the Am1 position. For each substrate, the shortest distances to water molecules which come within the cutoff distance or 3 Å during simulation are shown. For NH_3/NH_4^+, the shortest H-bond distances are reported. The dashed line is the cutoff distance used for coordination number calculations.

A3

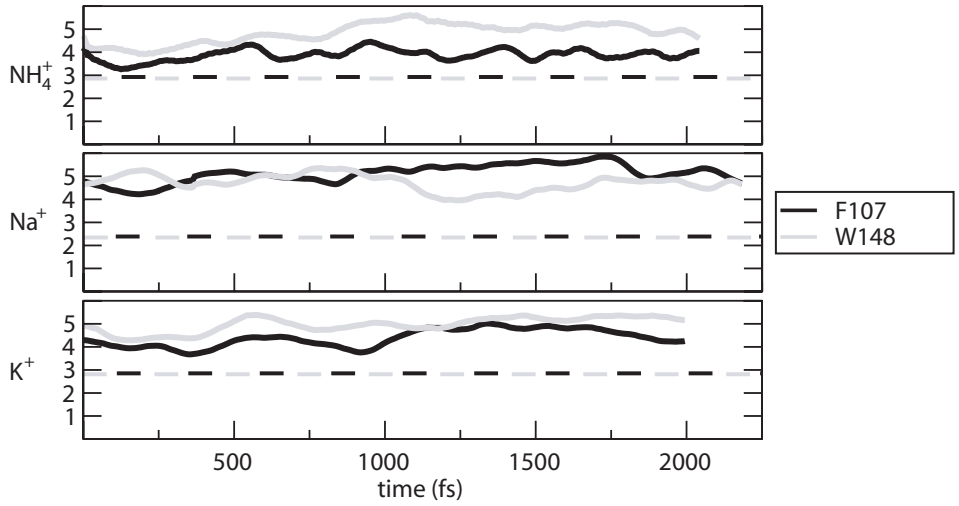

Figure A3: Cation-π distances for charged substrates at the Am1 position. For each charged substrate, the cation-π distances measured from the substrate center atom (N, K or Na) to the geometric center of the aromatic rings in F107 and W148 are shown. In the latter case, it is the distance to the five- or six-membered ring, whichever is shortest. The dashed lines are reference distances reported for optimized geometries of substrate interactions with aromatic motifs [193].

A4

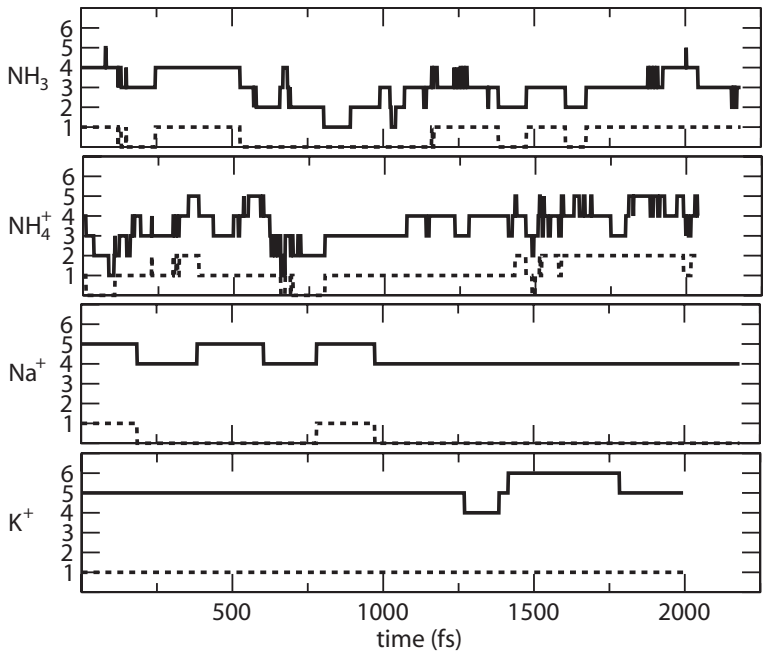

Figure A4: Coordination numbers for substrates at the Am1 position. Coordination numbers calculated using the distance cutoffs listed in Table 6.3 in the article. The dashed line is the contribution from the protein.

A5

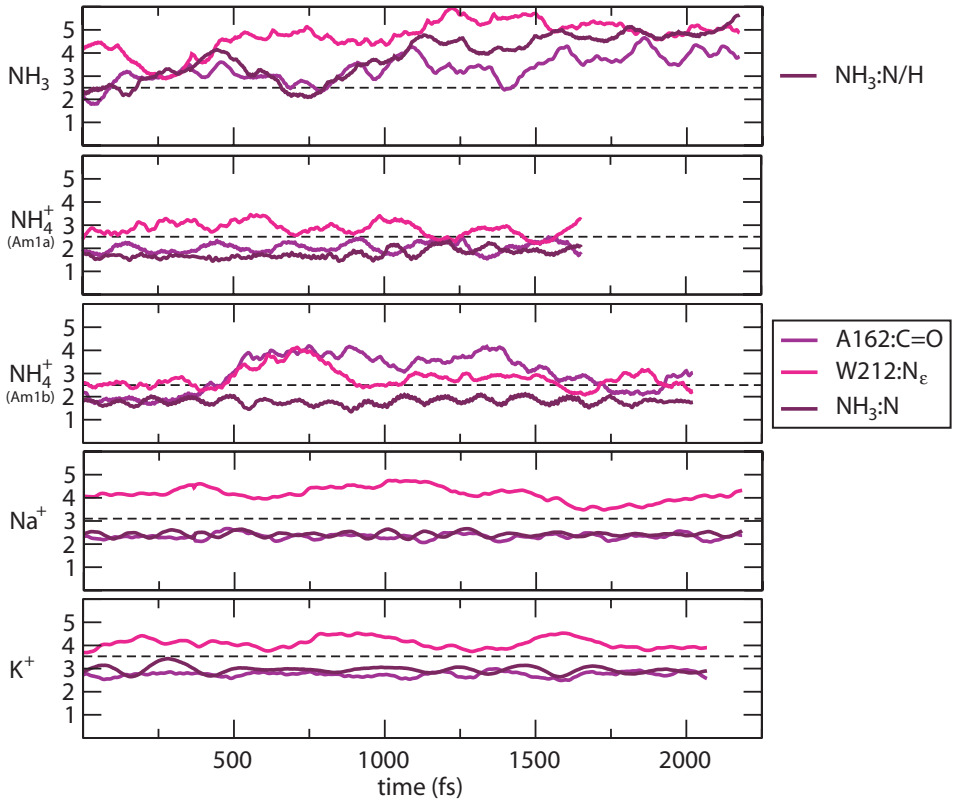

Figure A5: Distance between substrates and nearby residues at the Am1a and Am1b positions. For each substrate the shortest distances to residues A162 and W212, and to the closest intraluminal NH_3 molecule are shown. For NH_3/NH_4^+ the shortest H-bond distances are reported. The dashed line is the cutoff distance used for coordination number calculations.

A6

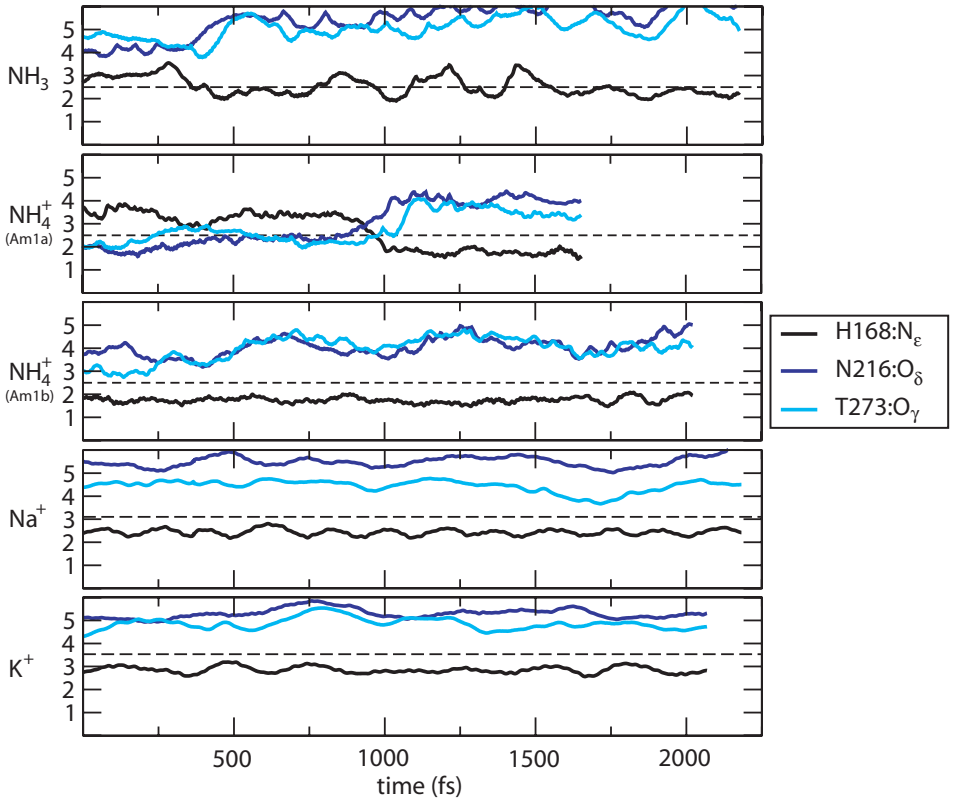

Figure A6: Distance between substrates and nearby residues at the Am1a and Am1b positions. For each substrate the shortest distances to residues H168, N216 and T273 are shown. For NH_3/NH_4^+ the shortest H-bond distances are reported. The dashed line is the cutoff distance used for coordination number calculations.

A7

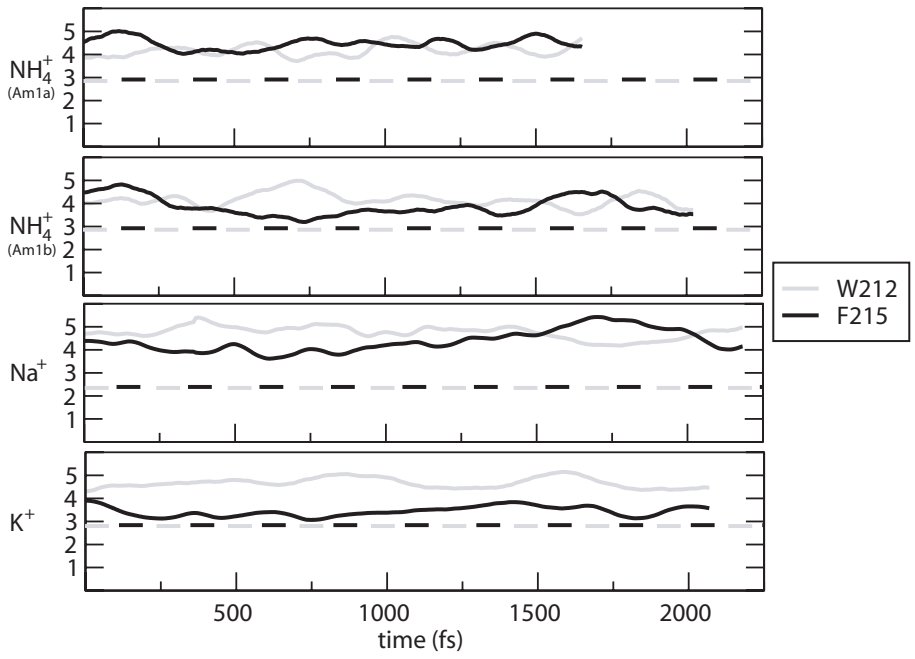

Figure A7: Cation-π distances for charged substrates at the Am1a and Am1b positions. For each charged substrate the cation-π distances measured from the substrate center atom (N, K or Na) to the geometric center of the aromatic rings in W212 and F215 are shown. In the latter case, it is the distance to the five- or six-membered ring, whichever is shortest. The dashed lines are reference distances reported for optimized geometries of substrate interactions with aromatic motifs [193].

A8

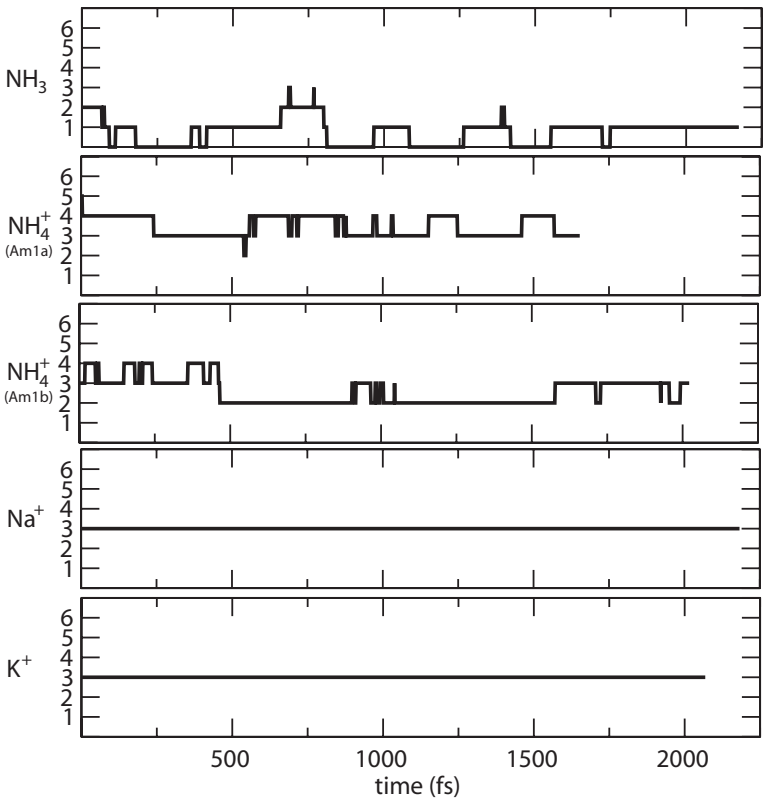

Figure A8: Coordination numbers for substrates at the Am1a and Am1b positions. Coordination numbers calculated using the distance cutoffs listed in Table 6.3 in the article.

A9

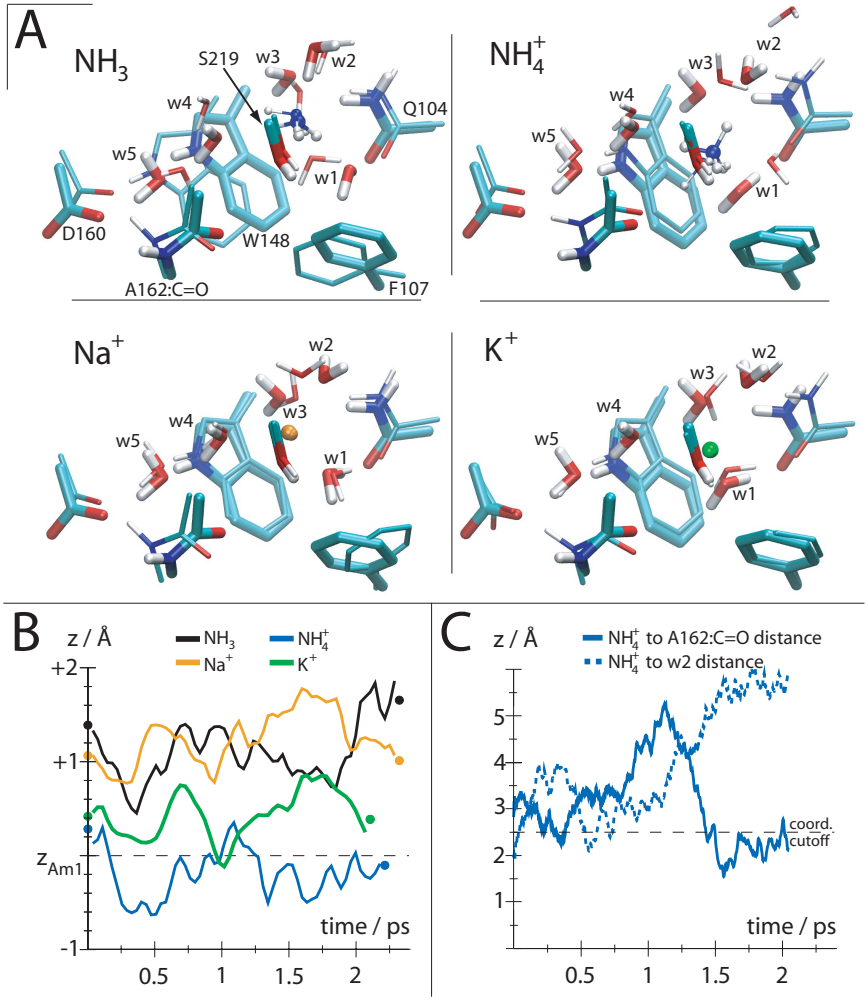

Figure A9: Substrates at the Am1 position. (A) Optimized structures with the four different substrates: NH_3, NH_4^+, Na^+ (*orange*), and K^+ (*green*), at the Am1 position. The initial structures (*thick licorice/ball-and-stick*) used as starting points for the simulations are overlayed with the final structures (*thin licorice/ball-and-stick*) after ∼2 ps simulation. All residues included in the cluster calculations in Table B3 are shown. In some cases, one of the water molecules might have been excluded from the cluster calculation although all five water molecules are shown on each figure. Distances between the substrate and coordinating residues are listed in Table B1. (B) z-positions of substrate heavy atoms relative to the Am1 position along the trajectories as well as the z-positions of the optimized structures (filled circles). (C) The distances between NH_4^+ and A162:C=O, and water molecule w2, along the trajectory reveal a correlation between the loss of a coordinating water and coordination to A162:C=O.

A10

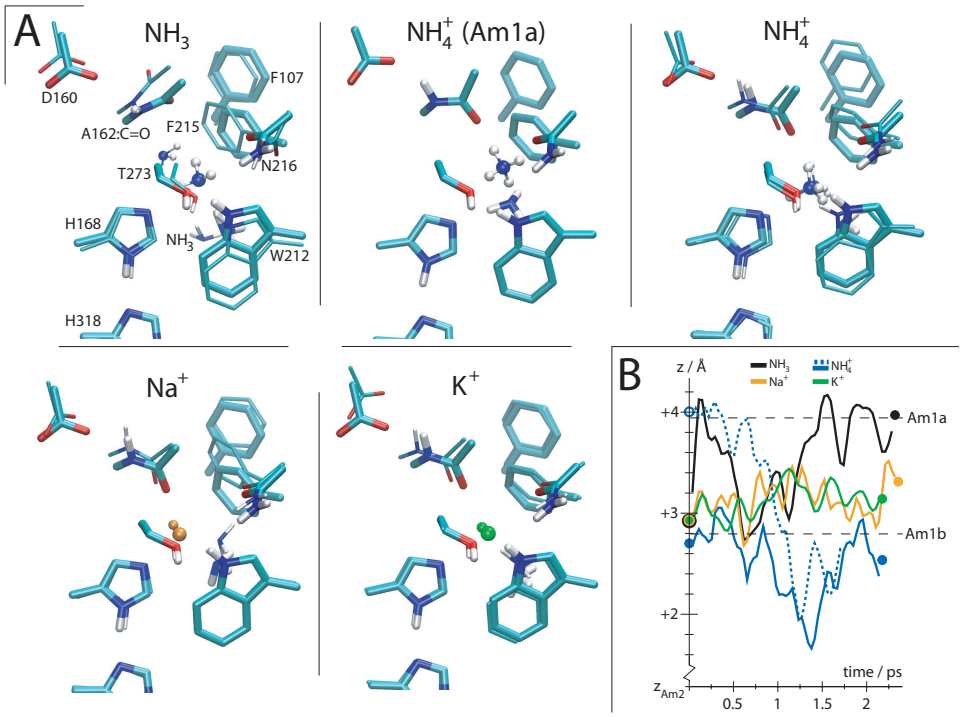

Figure A10: Substrates at the Am1a and Am1b positions. (A) Optimized structures with NH_4^+ at the Am1a position and the four different substrates: NH_3, NH_4^+, Na^+ (*orange*), and K^+ (*green*), at the Am1b position. The initial structures (*thick licorice/ball-and-stick*) used as starting points for the simulations are overlayed with the final structures (*thin licorice/ball-and-stick*) after ∼2 ps simulation. Only the initial structure of NH_4^+ at the Am1a is shown since it translocated to the Am1b position during simulation. All residues which were included in the cluster calculations in Table B3 are shown. Distances between the substrate and coordinating residues are listed in Table B2. (B) z-positions of substrate heavy atoms relative to the Am2 position along the trajectories as well as the z-positions of the optimized Am1a (empty circles) and Am1b structures (filled circles).

Appendix B

Supplementary Tables to Chapter 6

B1

	$\mathbf{NH_3}$[0/†]	$\mathbf{NH_4^+}$[0/†]	$\mathbf{Na^+}$[0/†]	$\mathbf{K^+}$[0/†]
Q104:O_ϵ	4.17/3.89	3.64/3.92	4.44/4.34	4.76/4.51
Q104:N_ϵ	3.59/4.23	3.61/4.83	3.97/4.09	4.07/4.74
W148:N_ϵ	4.37/6.10	3.78/3.95	4.64/4.62	4.87/5.09
D160:O_δ	7.05/6.56	6.44/5.98	7.37/7.37	7.42/7.48
A162:C=O	4.31/5.25	3.49/1.95	4.44/5.63	4.10/4.62
S219:O_γ	2.24/2.12	2.73/1.95	3.48/3.94	3.03/2.82
w1	1.97/2.13	1.58/1.66	2.26/2.24	2.63/2.67
w2	1.68/1.73	1.88/6.07	2.38/2.36	2.74/3.22
w3	3.21/3.45	2.79/2.47	2.43/2.38	2.91/2.88
w4	2.04/2.44	2.00/2.21	2.41/2.40	2.83/2.82
w5	4.56/3.93	4.03/4.04	4.86/4.83	5.02/5.04
F107	(4.28/4.71)	4.29/4.14	5.01/5.22	4.42/4.46
W148	(4.55/5.61)	4.67/4.91	4.62/4.46	5.02/5.20

Table B1: Coordination Distances at the Am1 position. Distances between coordinating residues/ water molecules and the four substrates in the optimized structures are listed. For NH_3/NH_4^+ the listed distances are the shortest H-bond distances, except for the last two entries, which are cation-π distances measured from the substrate center atom (N, K or Na) to the geometric center of the aromatic rings. All distances are in Ångströms. 0 and † denote initial and final structures, respectively.

B2

	$\mathbf{NH_3}$[0/†]	$\mathbf{NH_4^+}$[0/†]	$\mathbf{Na^+}$[0/†]	$\mathbf{K^+}$[0/†]
A162:C=O	3.34/4.00	(1.80) 1.97/2.49	2.32/2.29	2.67/2.79
H168:N_ϵ	2.16/2.12	(3.68) 1.75/1.67	2.39/2.39	2.81/2.80
W212:N_ϵ	3.33/5.06	(2.75) 2.71/2.46	3.97/4.29	3.57/3.86
N216:O_δ	4.44/6.12	(2.16) 4.01/4.70	5.41/5.71	4.88/5.35
N216:N_δ	6.25/7.11	(4.07) 5.65/6.17	7.20/7.56	6.71/7.15
T273:O_γ	4.40/5.14	(2.02) 3.20/3.95	4.37/4.65	4.24/4.64
NH_3	2.05/5.06	(1.61) 1.76/1.81	2.47/2.40	2.89/2.89
W212	(3.92/5.85)	(3.90) 4.15/4.01	4.57/4.93	4.13/4.41
F215	(2.88/3.02)	(4.10) 4.13/3.90	4.15/4.56	3.71/3.69

Table B2: Coordination Distances at the Am1a and Am1b positions. Distances between coordinating residues and the four substrates in the optimized structures are listed. The entries in parenthesis are for the the initial Am1a structure. For NH_3/NH_4^+ the listed distances are the shortest H-bond distances, except for the last two entries, which are cation-π distances measured from the substrate center atom (N, K or Na) to the geometric center of the aromatic rings. All distances are in Ångströms. 0 and † denote initial and final structures, respectively.

B3

Substrate:		NH_3		NH_4^+		Na^+		K^+	
Position	**Fragment**	E_{int}[0/†]	%[0/†]	E_{int}[0/†]	%[0/†]	E_{int}[0/†]	%[0/†]	E_{int}[0/†]	%[0/†]
Am1	Q104	-2.7/-2.2	12/10	-5.3/-8.4	4/6	-0.2/-1.8	0/1	-0.3/-4.3	0/3
	F107	0.4/0.7	-2/-3	-2.0/-3.3	1/2	-3.5/-4.1	2/3	-1.8/-2.6	1/2
	W148	0.2/-0.2	-1/1	-3.6/0.9	3/-1	-3.1/-4.6	2/3	-3.5/-2.8	3/2
	D160	-2.6/-2.7	12/13	-45.4/-49.9	33/35	-46.1/-46.7	30/30	-45.2/-46.2	34/36
	A162:C=O	-0.7/0.1	3/-1	-7.1/-16.9	5/12	-6.0/-1.8	4/1	-6.9/-5.1	5/4
	S219	-1.9/-2.1	9/10	-5.7/-10.1	4/7	-5.4/-5.3	3/3	-8.3/-10.3	6/8
	w1	-2.5/-3.5	11/16	-20.4/-16.5	15/11	-21.6/-22.0	14/14	-15.1/-15.3	11/12
	w2	-12.3/-12.0	57/56	-10.5/NI	8/NI	-13.5/-13.7	9/9	-9.9/-4.2	7/3
	w3	NI/NI	NI/NI	NI/-4.7	NI/3	-12.8/-12.9	8/8	-6.0/-4.4	4/3
	w4	-2.5/-1.6	12/8	-13.3/-9.3	10/6	-18.8/-19.6	12/13	-13.3/-14.0	10/11
	w5	-1.3/-1.1	6/5	-6.0/-6.8	4/5	-7.5/-9.1	5/6	-6.9/-7.7	5/6
	ΔE_{int}(ALL)	-21.5/-21.3	100/100	-136.0/-144.1	100/100	-154.0/-153.8	100/100	-132.4/-128.0	100/100
(Am1a) Am1b	A162:C=O	-0.3/-0.4	4/7	(-13.9) -13.6/-11.2	(15) 15/12	-23.1/-25.0	22/24	-16.6/-16.4	19/19
	H168	-3.5/-2.9	47/54	(-10.7) -20.4/-24.1	(12) 22/26	-25.2/-26.5	24/25	-17.4/-18.7	20/22
	W212	1.2/0.0	-16/0	(5.5) -3.5/-5.1	(-6) 4/5	-2.1/-2.1	2/2	-3.1/-2.7	4/3
	F215	0.1/-1.1	-2/21	(-2.8) -3.5/-4.6	(3) 4/5	-4.4/-3.8	4/4	-5.6/-6.0	7/7
	N216	-0.3/-0.1	4/2	(-13.4) -5.6/-4.5	(15) 6/5	-5.2/-4.8	5/5	-6.1/-5.4	7/6
	T273	0.4/0.2	-6/-4	(-7.0) -1.0/-2.1	(8) 1/2	-1.2/-1.1	1/1	-2.0/-1.6	2/2
	NH3	-3.4/0.1	45/-2	(-22.4) -17.0/-16.6	(24) 18/18	-20.2/-21.2	19/20	-12.0/-13.4	14/16
	ΔE_{int}(ALL)	-7.5/-5.3	100/100	(-91.7) -93.0/-93.4	(100) 100/100	-105.4/-105.7	100/100	-85.3/-84.2	100/100

Table B3: Interaction Energies. Interaction energies between substrates and coordinating residues obtained from cluster calculations. All energies are in kcal/mol. Also listed are the percentage each interaction energy comprise of the total interaction energy ΔE_{int}(ALL). Since they are not pairwise additive they do not sum up to a total of 100%, but to ∼90% for the charged substrates and ∼120% for NH_3. A negative percentage arise for repulsive interactions. 0 and † denote initial and final structures, respectively. Entries in parenthesis are the energies of NH_4^+ initially at the Am1a position.